PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-041345
(43)Date of publication of application : 08.02.2002

(51)Int.Cl. G06F 12/00
G06F 11/20
G06F 12/16

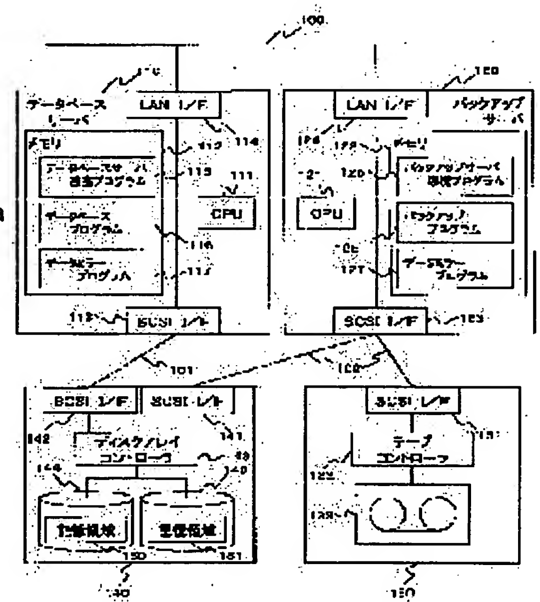
(21)Application number : 2000-231391
(22)Date of filing : 27.07.2000

(71)Applicant : HITACHI LTD
(72)Inventor : AJIMATSU YASUYUKI
MATSUNAMI NAOTO
MIZUNO YOICHI

(54) COMPUTER SYSTEM AND ITS DATA BACKUP METHOD

(57)Abstract:
PROBLEM TO BE SOLVED: To perform high speed backup by preventing the lowering of performance of an online processing during a backup processing.
SOLUTION: Data write from application 116 to an external storage device 160 is duplexed in storage areas 150, 151 by a middleware 117 on a server 110. A duplex operation is stopped in the case of a backup operation, a processing of the application is continued by using one storage area 150 and backup data are transferred from the other storage area 151 to a tape device 130 by a backup program 126 of a backup server 120 in the case of the backup. The duplex operation is restarted by the middleware 117 and the contents of two storage areas are resynchronized after the completion of the backup.

図 1



LEGAL STATUS
[Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's

decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　　(12) 公開特許公報（Ａ）　　(11)特許出願公開番号
特開2002－41345
（P2002－41345A）
(43)公開日　平成14年2月8日(2002.2.8)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード゛(参考)
ＧＯ６Ｆ　12/00	５３１	ＧＯ６Ｆ　12/00	５３１Ｍ	５Ｂ０１８
11/20	３１０	11/20	３１０Ａ	５Ｂ０３４
12/16	３１０	12/16	３１０Ｍ	５Ｂ０８２

審査請求　未請求　請求項の数８　ＯＬ　（全 10 頁）

(21)出願番号　特願2000－231391(P2000－231391)

(22)出願日　平成12年7月27日(2000.7.27)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地

(72)発明者　味松　康行
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　松並　直人
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

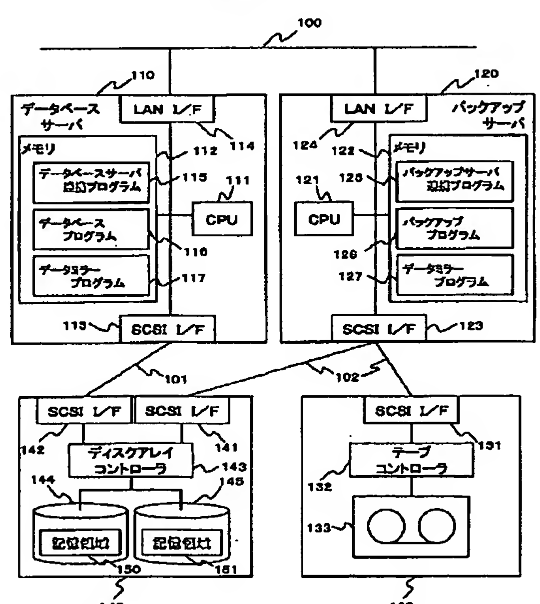
(74)代理人　100075096
弁理士　作田　康夫

最終頁に続く

(54)【発明の名称】　コンピュータシステム及びそのデータバックアップ方法

(57)【要約】
【課題】バックアップ処理中におけるオンライン処理の性能低下を防止し、高速なバックアップを行えるようにする。
【解決手段】アプリケーション１１６から外部記憶装置１６０へのデータ書き込みを、サーバ１１０上のミドルウェア１１７で記憶領域１５０、１５１に二重化する。バックアップ時は二重化を停止し、一方の記憶領域１５０を使ってアプリケーションの処理を継続し、バックアップサーバ１２０のバックアッププログラム１２６により他方の記憶領域１５１からテープ装置１３０にバックアップデータを転送する。バックアップ終了後、ミドルウェア１１７により二重化を再開すると共に、２つの記憶領域の内容を再同期させる。

図 1



【特許請求の範囲】
【請求項１】第１の記憶領域及び第２の記憶領域を有する第１の記憶装置と、
前記記憶装置に保持されたデータを処理するアプリケーションプログラムと、該アプリケーションプログラムによるデータの書き込みを前記第１の記憶領域と前記第２の記憶領域に二重化する二重化手段とを有する第１のコンピュータと、
前記第１の記憶装置に保持されたデータを第２の記憶装置にコピーするバックアップ手段を有する第２のコンピュータとを備え、
前記第１のコンピュータは、さらに、前記バックアップ手段からの要求に応答して、前記二重化手段により前記第１の記憶領域と前記第２の記憶領域との間の二重化を停止させ、あるいは、二重化が停止されている前記第１の記憶領域と前記第２の記憶領域の二重化を前記二重化手段に再開させる連携手段を有し、前記バックアップ手段は、前記第１の記憶領域と前記第２の記憶領域の二重化が停止されている間に、前記第２の記憶領域に保持されたデータを前記第２の記憶装置に転送することを特徴とするコンピュータシステム。
【請求項２】前記二重化手段は、二重化を停止中に更新された前記第１の記憶領域の位置を記録する手段を有し、二重化の再開時に、前記記録手段に記録された更新位置に基づいて更新された前記第１の記憶領域の位置に保持されたデータを前記第２の記憶領域にコピーすることを特徴とする請求項１記載のコンピュータシステム。
【請求項３】前記記憶装置は、さらに第３の記憶領域を有し、前記第１のコンピュータは、前記二重化状態が停止された前記第２の記憶領域から前記第３の記憶領域にデータをコピーする手段を有し、前記バックアップ手段は、前記第２の記憶領域に替えて前記第３の記憶領域に保持されたデータを前記第２の記憶装置に転送することを特徴とする請求項２記載のコンピュータシステム。
【請求項４】前記連携手段は、前記二重化手段により前記第１の記憶領域と前記第２の記憶領域の二重化状態を停止した後、前記第２の記憶領域を書き込み禁止の属性に設定し、前記バックアップ手段によるデータの転送が終了した後、前記第２の記憶領域に対する書き込み禁止の属性を解除することを特徴とする請求項１記載のコンピュータシステム。
【請求項５】前記第１の記憶装置は、アクセス元のコンピュータに割り当てられた識別情報に基づいて、前記第１及び前記第２の記憶領域へのアクセスの可否を制御する手段を有しており、前記連携手段は、通常の稼働時において前記第２のコンピュータからのアクセスが禁止される前記第２の記憶領域を、前記二重化状態を停止した後、前記第２のコンピュータからアクセス可能とするように前記記憶装置に指示し、前記バックアップ手段によるデータの転送が終了した後、前記第２の記憶領域に対する前記第２のコンピュータからアクセスを禁止するように指示することを特徴とする請求項１記載のコンピュータシステム。
【請求項６】前記第１のコンピュータ及び前記第２のコンピュータと前記第１の記憶装置との間に配置され、アクセス元のコンピュータに割り当てられた識別情報に基づいて、前記第１及び前記第２の記憶領域へのアクセスの可否を制御する手段を備える接続装置を有し、前記連携手段は、通常の稼働時において前記第２のコンピュータからのアクセスが禁止される前記第２の記憶領域を、前記二重化状態を停止した後、前記第２のコンピュータからアクセス可能とするように前記接続装置に指示し、前記バックアップ手段によるデータの転送が終了した後、前記第２の記憶領域に対する前記第２のコンピュータからアクセスを禁止するように指示することを特徴とする請求項１記載のコンピュータシステム。
【請求項７】アプリケーションプログラムが動作する第１のコンピュータ、第２のコンピュータ、前記第１及び前記第２のコンピュータと接続された第１の記憶装置、及び前記第２のコンピュータに接続された第２の記憶装置を有するコンピュータシステムにおけるデータのバックアップ方法において、
前記アプリケーションプログラムによるデータの書き込み要求に応じて、該書き込み要求に基づく書き込みデータを前記第１の記憶装置の第１の記憶領域及び第２の記憶領域に二重化して書き込み、
前記第１のコンピュータにより前記第１の記憶領域と前記第２の記憶領域との間の二重化を停止し、
前記第２のコンピュータにより前記第２の記憶領域に保持されたデータを前記第２の記憶装置に転送し、
前記第１のコンピュータにより、前記第１の記憶領域の内容と前記第２の記憶領域の内容とを一致させ、前記第１の記憶領域と前記第２の記憶領域を二重化状態に戻すことを特徴とするバックアップ方法。
【請求項８】前記第２のコンピュータから前記第２の記憶領域へのアクセスは通常のシステム稼働時には禁止されており、前記二重化を停止するステップは、前記第２の記憶領域への前記第２のコンピュータからのアクセスを可能とするステップを含み、前記二重化状態に戻すステップは、前記第２のコンピュータから前記第２の記憶領域へのアクセスを禁止状態に戻すステップを含むことを特徴とする請求項７記載のバックアップ方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、コンピュータシステム及びそのデータのバックアップ方法に係り、特に、通常業務の運用を停止することなくデータのバックアップを行うのに好適なコンピュータシステム及びそのデータのバックアップ方法に関する。
【０００２】

【従来の技術】一般にコンピュータシステムでは、装置の障害、ソフトウェアの欠陥、誤操作などにより外部記憶装置に記録されたデータが喪失した場合に備え、磁気テープなどの記憶媒体にデータをコピーして保存すること(バックアップ)が行われている。

【0003】データのバックアップ処理中にアプリケーションプログラムを停止することができない場合、アプリケーションプログラムのバックアップ支援機能が利用される。バックアッププログラムは、アプリケーションプログラムと同じコンピュータ上で動作し、アプリケーションプログラムと同じデータにアクセスする。バックアップ処理中は、アプリケーションプログラムのバックアップ支援機能により、バックアップデータの整合性が失われないようにされる。

【0004】例えば、データベースをバックアップする場合、データベースシステムは、バックアップ中も処理を継続する。バックアップ中、データベースシステムは、バックアップデータの書き換えを抑制し、あるいは、書き換え途中のデータがバックアップされてもバックアップデータの整合性が失われないようにログに補償情報を残す。このようなデータベースシステムの機能を利用することにより、データベース処理を停止することなく整合性のあるデータのバックアップを取得することができる。

【0005】

【発明が解決しようとする課題】上述した従来技術によれば、データベース処理などのオンライン処理を行うアプリケーションプログラムとバックアップ処理を実行するバックアッププログラムが同じコンピュータ上で実行される。また、両者は、同じデータにアクセスするため、コンピュータのCPU、およびディスクシステムなどの外部記憶装置の負荷が高まり、オンライン処理およびバックアップ処理の性能が低下するといった問題があった。

【0006】したがって、本発明の目的は、バックアップ処理中におけるオンライン処理の性能低下を防止し、高速なバックアップを行えるようにすることにある。

【0007】

【課題を解決するための手段】上記目的を達成するために、本発明によるコンピュータシステムは、第1の記憶領域及び第2の記憶領域を有する第1の記憶装置と、記憶装置に保持されたデータを処理するアプリケーションプログラムと、該アプリケーションプログラムによるデータの書き込みを第1の記憶領域と第2の記憶領域に二重化する二重化手段とを有する第1のコンピュータと、第1の記憶装置に保持されたデータを第2の記憶装置にコピーするバックアップ手段を有する第2のコンピュータとを備えて構成される。

【0008】本発明の好ましい態様において、第1のコンピュータは、さらに、バックアップ手段からの要求に応答して、二重化手段により前記第1の記憶領域と前記第2の記憶領域との間の二重化を停止させ、あるいは、二重化が停止されている第1の記憶領域と第2の記憶領域の二重化を二重化手段に再開させる連携手段を有する。バックアップ手段は、第1の記憶領域と第2の記憶領域の二重化が停止されている間に、第2の記憶領域に保持されたデータを前記第2の記憶装置に転送する。

【0009】本発明のより好ましい態様において、二重化手段は、二重化を停止中に更新された第1の記憶領域の位置を記録する手段を有し、二重化の再開時に、記録手段に記録された更新位置に基づいて更新された第1の記憶領域の位置に保持されたデータを第2の記憶領域にコピーする。

【0010】また、本発明の他の態様において、第1の記憶装置は、第3の記憶領域を備え、第1のコンピュータは、二重化手段により第1の記憶領域と第2の記憶領域の間の二重化が停止されている間に、第2の記憶領域から第3の記憶領域にデータを転送する手段を有する。バックアッププログラムは、第3の記憶領域からデータを読み出し、第2の記憶装置に転送する。

【0011】

【発明の実施の形態】図1は、第1の実施形態によるコンピュータシステムの構成を示す簡略なブロック図である。本実施形態におけるコンピュータシステムでは、データベースプログラムがアクセスするデータが、データベース処理を継続したままバックアップ(オンラインバックアップ)される。

【0012】図に示すように、本実施形態におけるコンピュータシステムでは、データベースサーバ110とバックアップサーバ120が、それぞれLANインタフェース114、124を介してLAN100で接続される。データベースサーバ110は、インタフェース113、バス101を介してディスクアレイ140のインタフェース142に接続される。バックアップサーバ120は、インタフェース123、バス102を介してディスクアレイ140のインタフェース141と、テープ装置130のインタフェース131に接続される。

【0013】データベースサーバ110は、メモリ112内にCPU111により実行されるデータベースプログラム116、データミラープログラム117、およびデータベース連携プログラム115を有する。バックアップサーバ120は、メモリ122内にCPU121によって実行されるバックアッププログラム126、データミラープログラム127、およびバックアップサーバ連携プログラム125を有する。

【0014】ディスクアレイ140は、データベースサーバ110、及びバックアップサーバ120によって共有される外部記憶装置である。ディスクアレイ140は、2つのディスクドライブ144、145内に記憶領域150、151を持ち、ディスクアレイコントローラ

143によってその動作が制御される。バックアップ装置として使用されるテープ装置130は、テープドライブ133を持ち、テープコントローラ132によりその動作が制御される。記憶領域150、151は同じ外部記憶装置内にある必要はなく、それぞれ独立した個別の単体ディスク装置に存在してもよい。ディスクアレイ140の記憶領域150は、データベースサーバ110からアクセスできるようマウントされている。

【0015】データベースプログラム116は、記憶領域150に格納されたデータベースに対する読み書きを行う。データベースプログラム116は、ユーザや他のプログラムからの指示により、バックアップモードで動作する。バックアップモードでは、データベースプログラム116は、データベース処理は継続するがバックアップ中の外部記憶150への書き込みは抑制するように動作する。バックアップモード中の記憶領域150に対する更新内容は、メモリ112内に設けられたバッファや、他の外部記憶領域に蓄積され、バックアップモードの終了時に記憶領域150に反映される。

【0016】データミラープログラム117、127は、例えば、デバイスドライバのようなミドルウェアとして動作する。データミラープログラム117、127は、記憶領域150と記憶領域151を二重化して管理し、データベースプログラム116など、他のプログラムによる記憶領域150への書き込みデータを記憶領域151にも書き込む“書き込み二重化機能”を有する。本明細書では、書き込み二重化機能により二重化された記憶領域のペアを単に、「ミラー」と呼ぶことにする。

【0017】データミラープログラム117、127は、また、ユーザや他のプログラムからの指示により二重化を停止し、記憶領域151を独立の記憶領域として上位装置に提供する“ミラー分離機能”、ユーザや他のプログラムからの指示により、記憶領域150の内容を記憶領域151にコピーし、2つの記憶領域が同一の内容を持つように再同期させる“ミラー再同期機能”を有する。ミラー再同期機能による再同期処理中に、記憶領域150への書き込みが発生すると、書き込み二重化機能により、同一のデータが記憶領域151に書き込まれて二重化される。

【0018】バックアッププログラム126は、ユーザからのバックアップ開始の指示やバックアップ対象の指定を受け付けるための“ユーザインタフェース機能”、ディスクアレイの記憶領域からテープ装置130のテープにデータを転送するための“データ転送機能”、及び、データ転送のタイミングと連動し、転送処理の前後に他のプログラムを起動するための“連携プログラム呼出機能”を有する。

【0019】バックアップサーバ連携プログラム125は、バックアッププログラム126の連携プログラム呼出機能によって起動され、通信機能によりLAN100を介してデータベースサーバ連携プログラム115を起動する。バックアップサーバ連携プログラム125は、ディスクアレイ140の記憶領域151をマウント・アンマウントするための“ディスクボリュームマウント機能”を有する。

【0020】データベースサーバ連携プログラム115は、データベースプログラム116に対してバックアップモードで動作するよう指示するための“データベース制御機能”、データミラープログラムに対して、ミラーの分離・再結合を指示するための“ミラー分離/再結合指示機能”、及び、ディスクアレイ140の記憶領域150をマウント・アンマウントするための“ディスクボリュームマウント機能”を有する。データベースサーバ連携プログラム115は、処理が完了すると通信機能によりLAN100を介してバックアップサーバ連携プログラム125に処理の完了を報告する。

【0021】図2は、データベースプログラムの稼働中にデータベースのバックアップを取得する際に実施される処理の流れを示すフローチャートである。

【0022】バックアップを行う際、ユーザは、バックアップサーバ120操作して、バックアッププログラム126を起動し、バックアップの開始を指示する（ステップ400）。バックアッププログラム126は、連携プログラム呼出機能により、あらかじめ登録されているバックアップサーバ連携プログラム125を起動する（ステップ401）。バックアップサーバ連携プログラム125は、LAN100を介してデータベースサーバ110と通信してデータベースサーバ連携プログラム115を起動し、データベースサーバ連携プログラムからの処理の完了報告を待つ（ステップ402）。

【0023】データベースサーバ連携プログラム115は、まずデータベースプログラム116に対してバックアップモードへの移行を指示する。この指示に応答して、データベースプログラムは、記憶領域への書き込みを抑制する。これにより、記憶領域内のデータは整合性のある状態に保たれる（ステップ403）。

【0024】次に、データベースサーバ連携プログラム115は、データミラープログラム117に対してミラーの分離を指示する。これにより、以後の記憶領域150への更新は記憶領域151に反映されず、記憶領域151は独立した記憶領域となる（ステップ404）。データベースサーバ連携プログラム115は、ミラー分離後、再びデータベースプログラム116に対してバックアップモードの終了を指示する。この指示に応じて、データベースプログラムは記憶領域150への更新を再開する（ステップ405）。この後、データベースサーバ連携プログラム115は、処理の完了をバックアップサーバ連携プログラムに報告し、処理を終了する（ステップ406）。

【0025】なお、ここまでの処理において、バックア

ップモードに移行するデータベースやミラーを分離する記憶領域等の情報は、データベースサーバ連携プログラム内であらかじめ指定するか、通信機能によりバックアップサーバ連携プログラムから与えられる。

【0026】バックアップサーバ連携プログラム125は、データベースサーバ110から終了報告を受けると、あらかじめ指定された記憶領域151をバックアップサーバ120上のアプリケーションプログラムからアクセスできるようにマウントする（ステップ407）。記憶領域151をマウントすると、バックアップサーバ連携プログラム125は、処理を終了して制御をバックアッププログラム126に戻す（ステップ408）。

【0027】バックアッププログラム126は、制御が戻ると、記憶領域151からデータを読み出し、そのデータをテープ装置130に転送する（ステップ409）。テープ装置130へのデータ転送完了後、バックアッププログラム126は、再びバックアップサーバ連携プログラム125を起動する（ステップ410）。バックアップサーバ連携プログラム125は、記憶領域151をアンマウントした後（ステップ411）、データベースサーバ連携プログラム115を起動する（ステップ412）。

【0028】データベースサーバ連携プログラム115は、データミラープログラム117に対して記憶領域150と151の再同期を指示する（ステップ413）。この指示に応じて、データミラープログラムは、記憶領域150と1512が同じ内容を持つように、記憶領域150から記憶領域151にデータをコピーする。その後、データベースサーバ連携プログラム115およびバックアップサーバ連携プログラム125は、それぞれの処理を終了してバックアッププログラム126に制御を戻し、バックアップ処理が完了する（ステップ414，415）。

【0029】以上の処理により、データベースプログラムの稼働中にバックアップを取得し、その処理を自動化することができる。

【0030】図3は、バックアップしたデータをディスクアレイに書き戻すリストア処理のフローチャートである。

【0031】リストア処理は、データベースプログラム115が停止している状態で行われる。リストア処理の起動からデータベースサーバ連携プログラムの起動までの処理（ステップ500〜502）は、図2に示したバックアップ取得の処理におけるステップ400〜402で実施される処理とほぼ同様である。ただし、ステップ500において、ユーザは、バックアップの開始に替えてリストアの開始を指示する。

【0032】データベースサーバ連携プログラム115は、データベースサーバ110からディスクアレイ140の記憶領域150をアンマウントする。この処理により、リストア中にデータベースサーバ110から記憶領域150に対するアクセスが発生することを防ぐ（ステップ503）。記憶領域150をアンマウントすると、データベースサーバ連携プログラム115は、バックアップサーバ連携プログラム125に処理の完了を報告して処理を終了する（ステップ504）。

【0033】バックアップサーバ連携プログラム125は、バックアップサーバ120上のアプリケーションプログラムから記憶領域150をアクセスできるように、記憶領域150をマウントする（ステップ505）。この後、バックアップサーバ連携プログラム125は、処理を終了して制御をバックアッププログラム126に戻す（ステップ506）。処理が戻ると、バックアッププログラム126は、バックアップデータをテープ装置130から記憶領域150にリストアする。このとき、記憶領域150へのデータ書き込みは、データミラープログラム127により記憶領域151に二重化される（ステップ507）。

【0034】テープからのデータ転送（リストア）が終了した後の処理（ステップ508〜514）は、図2に示したバックアップ取得処理のステップ408〜414と同様に行われる。ただし、ここでは、バックアップ取得処理のステップ413における再同期指示に替えて、ステップ511において、データベース連携プログラム115により、記憶領域150を再びデータベースサーバ110にマウントする。

【0035】以上の処理により、バックアップしたデータをディスクアレイ140に元の状態でリストアすることができる。

【0036】本実施形態によれば、データベースプログラムとバックアッププログラムは、それぞれ異なるサーバ上で実行される。このため、データベース処理を行っているデータベースサーバのバックアップ処理によるCPU負荷の増大を防ぐことができる。また、各プログラムが処理のためにアクセスする記憶領域が異なるため、記憶領域へのアクセス負荷の集中を防止することもできる。

【0037】図4は、本発明の第2の実施形態におけるコンピュータシステムの構成を示す簡略なブロック図である。

【0038】本実施形態では、第一実施形態と同様にデータベースのオンラインバックアップを取得するが、データミラープログラムによる再同期処理がより高速化されている。また、本実施形態では、リストア処理において、バックアップサーバ内のデータミラープログラムは用いられない。以下、第1の実施形態と重複する部分については説明を省略し、主として第1の実施形態と異なる点について説明する。

【0039】図において、ディスクアレイ140は、さらに、第3のディスクドライブ146、およびディスク

ドライブ１４６に設けられる記憶領域１５３を有する。また、データベースサーバ１１０内のデータミラープログラム２１７は、差分再同期機能を持つ。ここで差分再同期機能とは、ミラーを分離して書き込みの二重化を停止している期間に更新された記憶領域１５１の位置をメモリ内に記録しておき、記憶領域１５１の内容を１５０に再同期させる際に、更新された記憶領域１５０内の位置の内容のみを記憶領域１５１にコピーする機能である。これにより、コピーするデータの量が低減し、再同期処理が高速化される。また、データベースサーバ連携プログラム２１５は、第１の実施形態におけるデータベースサーバ連携プログラム１１５の有する機能に加え、記憶領域１５１の内容を記憶領域１５２にコピーするためのデータコピー機能を備える。

【００４０】図5は、本実施形態におけるバックアップ処理の流れを示すフローチャートである。

【００４１】図において、ステップ６００～６０５では、図2に示したステップ４００～４０５と同様の処理が行われる。ミラーを分離した後、データベースサーバ連携プログラム２１５は、データコピー機能により、バックアップ対象のデータを記憶領域１５１から記憶領域１５２にコピーする（ステップ６０６）。記憶領域１５２へのコピー終了後、データベースサーバ連携プログラム２１５は、ミラープログラム２１７に対してミラーの差分再同期を指示する。この指示に応じて、ミラープログラムは、ミラー分離中に変更された記憶領域１５０内の部分を記憶領域１５１にコピーする。一般にディスク装置はテープ装置よりも高速にアクセスできる。したがって、記憶領域１５１から記憶領域１５２へのコピーは、テープ装置１３０へのコピーよりも短時間で終了する。ミラー分離時間が短縮されることにより、分離中に更新される記憶領域１５０内のデータ量が低減され、差分再同期処理はより短縮される（ステップ６０７）。

【００４２】ミラープログラム２１７への再同期指示後、データベースサーバ連携プログラム２１５は、処理の完了をバックアップサーバ連携プログラム１２５に報告し、処理を終了する（ステップ６０８）。ステップ６０９から６１２までは、図2に示したステップ４０７から４１０と同様の処理が実施される。ただし、第１の実施形態とは異なり、テープ装置１３０に転送されるバックアップデータは、記憶領域１５２から読み出される。また、データ転送終了後に起動されたバックアップサーバ連携プログラム１２５は、記憶領域１５３のアンマウント処理のみを行う（ステップ６１３）。以上の処理の後、バックアップサーバ連携プログラム２１５は処理を終了し、バックアップ処理が完了する（ステップ６１４）。

【００４３】本実施形態におけるリストア処理は、図3に示した第１の実施形態におけるリストア処理とほぼ同様にして行われる。ただし、本実施形態ではバックアップサーバ１２０がデータミラープログラムを持たないため、バックアッププログラム１２６によりリストアされるデータは、記憶領域１５０にのみ書き込まれる。このため、本実施形態では、ステップ５１１において記憶領域１５０のデータを記憶領域１５１にコピーする再同期処理が行われる。

【００４４】以上説明した第２の実施形態によれば、ミラー分離後、記憶領域１５１のデータは、一旦ディスクドライブ１４６の記憶領域１５２に転送され、記憶領域１５２からテープ装置に転送される。これにより、記憶領域１５０と記憶領域１５１の再同期処理は、記憶領域１５１から記憶領域１５２へのデータの転送が完了した時点で開始することができ、第１の実施形態と比較してミラー分離時間が短縮される。この結果、ミラー分離中に更新される記憶領域１５０のデータ量が低減され、再同期処理でコピーが必要とされるデータ量を低減することができる。また、再同期処理によるデータベースサーバやディスクアレイの負荷の増大する時間を短縮することができる。

【００４５】さらに、バックアップしたデータのリストア処理において、リストアするデータの二重化をデータベースサーバ側で行うため、バックアップサーバ側にデータミラープログラムが不要となる。

【００４６】図6は、本発明の第３の実施形態におけるコンピュータシステムの構成を示す簡略なブロック図である。

【００４７】ファイバチャネルで接続される既存のディスクアレイやファイバチャネルスイッチには、サーバーディスクアレイ間のアクセスをポート毎に制限する機能を持つものがある。また、サーバ上で動作するボリュームマネージャ等の既存のミドルウェアは、記憶領域（ボリューム）毎にアクセス属性を設け、「読み出し専用」とする機能を有するものがある。本実施形態では、バックアップ処理においてこれらの機能を利用し、記憶領域内のデータを誤ってバックアップサーバから変更することを防ぐようにしている。

【００４８】本実施形態において、データベースサーバ１１０、バックアップサーバ１２０、テープ装置１３０、及びディスクアレイ１４０は、それぞれファイバチャネルインタフェース３１３、３２３、３３１、３４１を備え、ファイバチャネルケーブル３６１～３６４を介してファイバチャネルハブ３７０に接続されている。ファイバチャネルハブ３７０は、データベースサーバ１１０、バックアップサーバ１２０、テープ装置１３０、及びディスクアレイ１４０相互の接続を制御し、これらの間でのデータ転送を実現する。

【００４９】ディスクアレイ１４０が有するディスクアレイコントローラ３４３は、データアクセス制御機能を持つ。データアクセス制御機能は、ファイバチャネルネットワークに接続されているワイルド・ワイド・ネーム

(World Wide Name) と呼ばれる各ポートの識別子に基づいて、記憶領域へのアクセスの許可・禁止を制御する。

【0050】本実施形態では、データベースサーバ連携プログラム315より、ファイバチャネルネットワークを通して設定情報を送ることにより、各ポートのアクセス許可・禁止を指定する。データベースサーバ連携プログラム315は、このための機能として、アクセス制御指示機能を有する。

【0051】記憶領域350、351は、それぞれ属性を記録する領域360、361を持つ。属性として読み出し専用属性が設定されている場合、サーバのミドルウェアは、その記憶領域に対するアプリケーションプログラムからの更新要求を拒否する。データベースサーバ連携プログラム315は、属性領域360、361に対して属性情報を書き込むことにより、各記憶領域の読み出し専用属性を指定・解除するボリューム属性操作機能を備える。

【0052】通常の稼動状態において、属性領域360、361には、読み出し専用属性は指定されず、記憶領域350、351は読み書き可能な状態とされる。また、バックアップサーバ120のファイバチャネルインタフェース323からのアクセスは、ディスクアクセスコントローラ343のデータアクセス制御機能により禁止される。

【0053】このほかの構成、各部の機能については、第1の実施形態と同様であり、ここでは説明を省略する。

【0054】本実施形態におけるバックアップ取得処理は、第1の実施形態とほぼ同様に実施される。

【0055】本実施形態では、ステップ404の後、データベースサーバ連携プログラム315のボリューム属性操作機能により、記憶領域351の内容が変更されないよう、属性領域361に読み出し専用の属性が指定される。データベースサーバ連携プログラム315は、さらに、データアクセス制御指示機能により、ディスクアレイコントローラ343に対してファイバチャネルインタフェース323から記憶領域351へのアクセスを許可するよう指示する。ファイバチャネルインタフェース323のWorld Wide Nameは、あらかじめ指定しておくか、あるいは、バックアップサーバ連携プログラム125から通信機能を使ってデータベースサーバ連携プログラム315に与えればよい。

【0056】バックアップサーバ連携プログラム125は、ステップ407において、属性領域361に読み出し専用の属性が設定されているため、記憶領域351を読み出し専用の領域としてマウントする。これにより、バックアップ処理中にバックアップサーバ120上のアプリケーションプログラムが誤って記憶領域351の内容を変更することを防止する。

【0057】ステップ413において再同期処理を開始する前に、データベースサーバ連携プログラム315は、データアクセス制御指示機能によりファイバチャネルインタフェース323からのアクセスを禁止するようディスクアレイコントローラ343に指示する。また、ボリューム属性操作機能により、属性領域361に設定されている読み出し専用の属性を解除し、記憶領域351を再び読み書き可能な状態とする。

【0058】その他の処理については、第1の実施形態と同様に実施される。したがって、ここでは、それらについての説明を省略する。

【0059】バックアップデータのリストア処理も第1の実施形態と同様に行われる。なお、リストア処理では、バックアップサーバ120から記憶領域にデータを書き込む必要がある。このため、リストア処理では、バックアップ取得処理とは異なりボリューム属性の変更は行われない。

【0060】本実施形態では、ファイバチャネルネットワークの構成にハブを用い、ディスクアレイにおいてアクセス制御を行っているが、ファイバチャネルスイッチを用いてスイッチ内でアクセス制御を行うこともできる。この場合には、あらかじめバックアップサーバおよびディスクアレイが接続されるポートを決定し、ポート間の通信を禁止・許可すればよい。

【0061】本実施形態によれば、通常の稼動状態においてバックアップサーバから記憶領域へのアクセスは禁止されるため、データベースが誤ってバックアップサーバから変更されることを防止できる。またバックアップ処理中も、アクセスする記憶領域のボリューム属性が読み出し専用に設定されるため、誤った更新を防止することができる。

【0062】

【発明の効果】以上述べたように、本発明によれば、オンライン処理とバックアップ処理が異なるサーバ上で実行され、オンライン処理とバックアップ処理がそれぞれ異なる記憶領域を使って行われる。この結果、バックアップ中のオンライン性能およびバックアップ性能の低下を防止できる。

【図面の簡単な説明】

【図1】第1の実施形態によるコンピュータシステムの構成を示す簡略なブロック図である。

【図2】データベースプログラムの稼働中にデータベースのバックアップを取得する際に実施される処理の流れを示すフローチャートである。

【図3】バックアップしたデータをディスクアレイに書き戻すリストア処理のフローチャートである。

【図4】第2の実施形態におけるコンピュータシステムの構成を示す簡略なブロック図である。

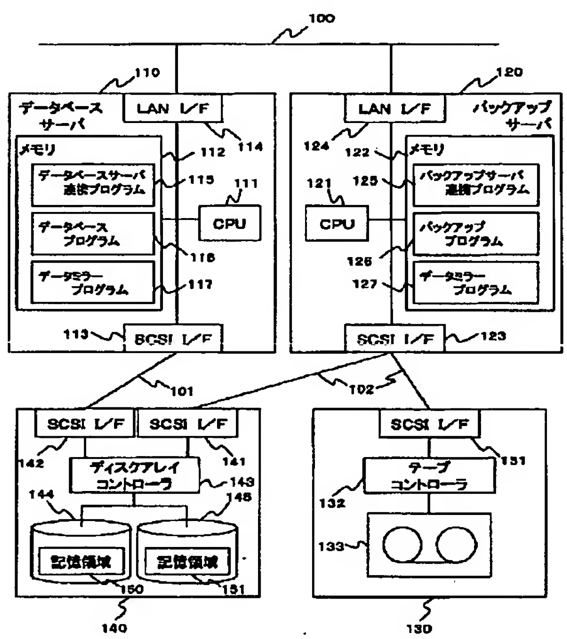
【図5】第2の実施形態におけるバックアップ処理の流れを示すフローチャートである。

【図6】第3の実施形態におけるコンピュータシステムの構成を示す簡略なブロック図である。

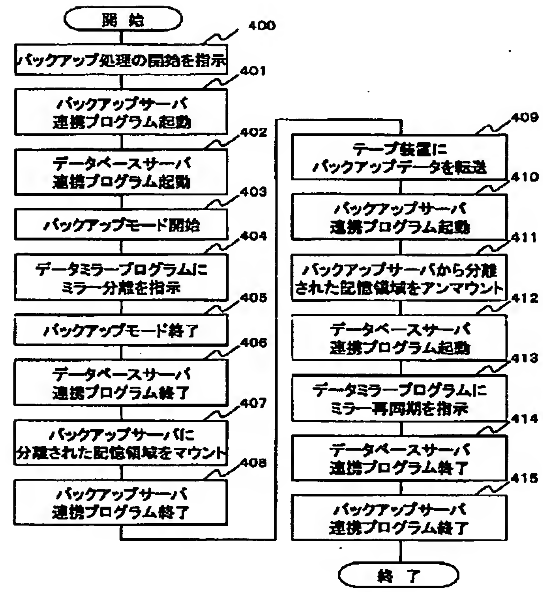
【符号の説明】

110…データベースサーバ
115…データベースサーバ連携プログラム
116…データベースプログラム
117…データミラープログラム
120…バックアップサーバ
125…バックアップサーバ連携プログラム
126…バックアッププログラム
127…データミラープログラム
130…テープ装置
140…ディスクアレイ

【図1】

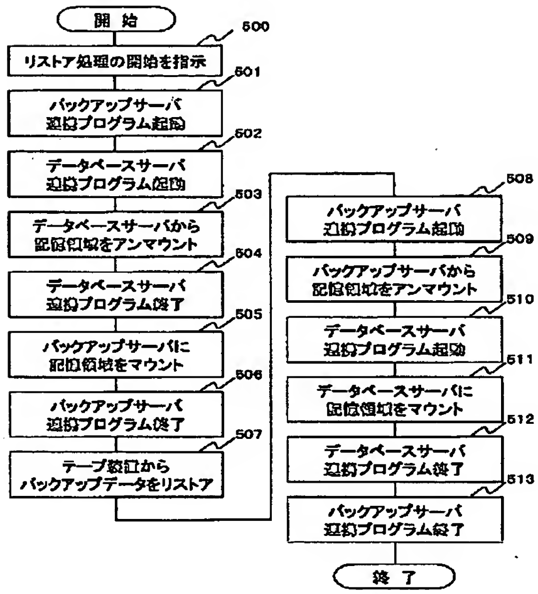


【図2】



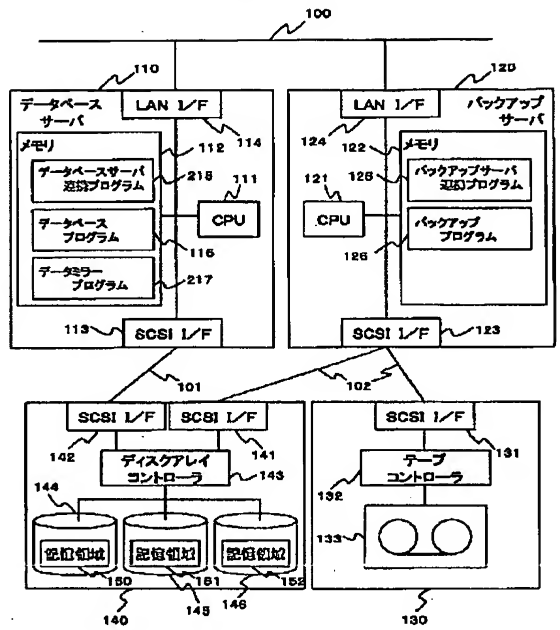
【図3】

図 3



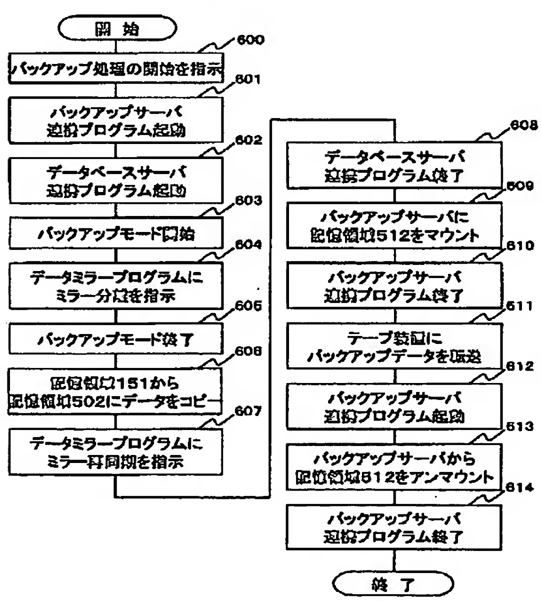
【図4】

図 4



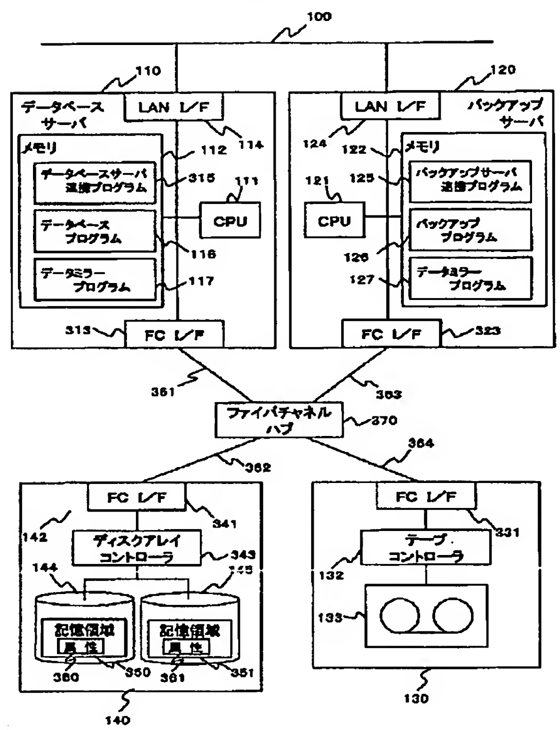
【図5】

図 5



【図6】

図 6



フロントページの続き

(72)発明者　水野　陽一
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

Fターム(参考)　5B018 GA04 HA04 MA12 QA20
5B034 AA01 BB17 CC04 DD05 DD06
5B082 DC12 DE04 DE06

US09/606050

[Name of document] Specification

[Title of the invention] Computer system using a storage area network and method of handling data in the computer system

[Background of the Invention]

The present invention relates to storage systems for storing data, in particular, a technique relating to methods for the data protection of handled data, the data sharing, the storage resource management, and the data handling.

At present, environment in which the information processing is performed has been changing drastically as a result of development of the Internet and Intranets, and expansion of such applications as data warehouse, electronic commerce, and information service, and this change has resulted in rapid increase in the amount of handled data.

For example, while the performance of CPUs has improved 100 times for the last five years, the input and output performance of disk drives has been held in about 10 times improvement. That is, the limit of the input and output performance compared with rapid increase in traffic has come to give rise to apprehensions. In addition, as applications such as enterprise resource planning (ERP), which processes a mass of data, and data warehouse have come to wide use, and information to be processed (documents, drawings, visual contents, etc.) has been diversified and communicated in Multimedia, demands of enterprises for a total disk capacity has increased two times a year on an average. Further, as storage capacities used in enterprises and others have increased and use of storages has been diversified, the

running cost of storages has also increased. Furthermore, backbone data in main frames has been shared and utilized by individual departments.

Described below is the situation of the information processing environment resulting from increase in the amount of handled data by using Fig. 2. As shown in Fig. 2, relations between servers and storages are established in such a way that, for example, a main frame (MF) as a server for a large-scale computer, a UNIX server as a server for a medium-scale computer, and a PC server as a server for a small-scale computer are connected with their respective exclusive storages, for example, RAIDs (Redundant Arrays of Inexpensive Disks) and magnetic tapes (MTs), and client computers give instructions to their respective servers via a LAN and perform data processing by using an exclusive storage for the relevant server.

Recently, proposed was a Storage Area Network (SAN) environment in which a SAN is constructed between the various servers and storages described above, and individual servers are allowed to access to any of the storages. Here, the SAN means a network that connects multiple servers and multiple storages through fiber channels, and is used only for input to and output from storages, and a SAN realizes the sharing of various storages, high-speed data processing between servers and storages, and long distance connection.

[Summary of the Invention]

As described above, an SAN is being introduced into environments, in which the information processing is performed, in order to improve the input and output performance, to expand a total disk capacity, to reduce the running cost of storages, and to expand data sharing. The SAN, as shown in Fig. 2, is a new type of

networks that connect multiple servers and multiple storages through a high-speed network (for example, fiber channels). In this environment, storages which are connected with their respective servers and are controlled by the servers are given independence from the servers, and at first a SAN used only for storages is constructed. In addition, all users that have an access right are enabled to share storage information on the SAN network.

In addition, connecting multiple storages enables to improve the input and output performance of the storages very significantly. That is, as merits, drastic improvement in the input and output performance of the storages (improvement in the performance), setting up and expanding flexibly a storage environment independently of server environments (improvement in scalability), unified storage operation (improvement in the storage management function), disaster measures by expanding the connection distance drastically (improvement in the data protection capability), etc. have been achieved.

However, existing proposals of SAN networks did not always disclose clearly concrete configurations or embodiments to realize these SAN network.

An object of the present invention is, in order to ensure the various merits and usability obtained by employing an SAN, to provide a integrated storage system in which collaboration over the entire storage system is reinforced by devising concrete functions of a storage system and corresponding concrete configurations, and in addition, another object is to provide a method for handling data more usefully at an Internet data center (abbreviated to "iDC"), which connects storages to the Internet and keeps and makes use of a large volume of data, by applying an integrated storage system to iDC.

In order to solve the issues described above, the present invention employs mainly the following configuration of a computer system and the following management method.

A computer system that is provided with multiple client computers, multiple various servers, multiple storages storing data, local area networks (LANs) connecting said computers and said servers, and a storage area network (SAN) lying between said servers and said storages, wherein said SAN forms circuit switched networks by fiber channel switches (FC switches) to make a mutual connection between any of said servers and any of said storages, and said SAN is equipped with terminals in which management and operation software has been installed to perform the storage management including management of logical volumes in said various storages, data arrangement, and error monitoring, the management of setup of said FC switches, and the data backup operation for data in said storages.

In addition, the management method is a method for managing a system comprising servers, storages storing data of said servers, and a network connecting said servers and said storages, and the method works in such a way that it obtains the information to identify data to be processed, obtains a specification of processing the data denoted by said information, gives said specification of processing to said storages keeping the data denoted by said information, and receives the result of processing the data denoted by said information from said storages.

[Brief Description of the Drawings]

Fig. 1 is a schematic diagram illustrating the basic overall configuration of an integrated storage system relating to a preferred embodiment of the present

invention.

Fig. 2 is a schematic diagram illustrating the overall configuration of a storage system according to a prior art.

Fig. 3 is a diagram describing the primary functions of an integrated storage system relating to a preferred embodiment of the present invention.

Fig. 4 is a diagram illustrating the basic system configuration about the non-disruptive backup in accordance with a preferred embodiment of the present invention.

Fig. 5a and Fig. 5b are a diagram describing functions or actions about the non-disruptive backup in accordance with a preferred embodiment of the present invention.

Fig. 6 is a diagram illustrating a system configuration in which mirroring software is used about the non-disruptive backup in accordance with a preferred embodiment of the present invention.

Fig. 7 is a diagram illustrating the preparations done in advance in a backup system and an example of system construction.

Fig. 8 is a diagram illustrating examples of various system configurations for backup by sharing tape units, relating to a preferred embodiment of the present invention.

Fig. 9 is a diagram illustrating a configuration for tape unit-shared backup in which multiple servers share one tape library.

Fig. 10 is a diagram illustrating a system configuration for asynchronous remote copying in disaster recovery, relating to a preferred embodiment of the present invention.

Fig. 11 is a diagram illustrating a system configuration for high-speed DB replication between servers in data sharing, relating to a preferred embodiment of the present invention.

Fig. 12 is a diagram illustrating error monitoring and backup operation in integrated system operation and management, relating to a preferred embodiment of the present invention.

Fig. 13 is a diagram illustrating centralized management of the storage performance in integrated system operation and management, relating to a preferred embodiment of the present invention.

Fig. 14 is a diagram illustrating storage management, in particular, the LUN manager and LUN security in integrated system operation and management, relating to a preferred embodiment of the present invention.

Fig. 15 is a diagram illustrating storage management, in particular, hierarchical control in a subsystem in integrated system operation and management, relating to a preferred embodiment of the present invention.

Fig. 16 is a diagram illustrating switch management; in particular, setting of zonings in integrated system operation and management, relating to a preferred embodiment of the present invention.

Fig. 17 is a diagram illustrating outline of a system configuration of an Internet data center in which an integrated storage system is used, relating to a preferred embodiment of the present invention.

Fig. 18 is a diagram illustrating storage integration in an Internet data center in accordance with a preferred embodiment of the present invention.

Fig. 19 is a diagram illustrating a system configuration for non-disruptive

backup in an Internet data center in accordance with a preferred embodiment of the present invention.

Fig. 20 is a diagram illustrating a system configuration for ensuring security in an Internet data center in accordance with a preferred embodiment of the present invention.

Fig. 21 is a diagram illustrating an example of system configurations of a large-scale computer system in which individual computer systems of multiple enterprises are connected mutually.

[Detailed Description of Preferred Embodiments]

The following describes a computer system in which a storage area network (SAN) is used and a method by which data is handled, referring to the drawings. Fig. 1 is a schematic diagram illustrating the basic overall configuration of said computer system relating to a preferred embodiment of the present invention.

In Fig. 1, the computer system in which the SAN is used consists of a main site and a remote site, and these sites are connected via a Wide Area Network (WAN). At the main site, multiple client computers and various servers, for example, a main frame (MF) as a server for large-scale computers, a UNIX server as a server for medium-scale computers, and a PC server as a server for small-scale computers, are connected via a LAN. In addition, a dedicated terminal in which operation and management software on integrated storage system has been installed is connected with the LAN, and the whole of the integrated storage system is operated, managed, and monitored by using the terminal. This operation and management software can be installed in any of the client terminals instead of the dedicated terminal and the

relevant client terminal is used for operation and management of the integrated storage system.

Further, storages such as a RAID, a tape library, and a DVD-RAM library/library array are connected with the server such as the main frame (MF) server, the UNIX server, and the PC server via a Storage Area Network (SAN) consisting of network switches such as a fiber channel switch (FC-Switch) and a fiber channel hub (FC-Hub) not shown in the figure. In addition, the main site is connected with the remote site consisting of the same components as those of the main site via a wide area communication network such as WAN.

Here, since the servers and the storages are connected through channel switches in the SAN, the servers and the storages which are connected through channel switches are enabled to be added, detached, and changed optionally. Therefore, firstly storages are enabled to be added and detached optionally to suit the storage capacity and the kind and object (access speed, cost, etc.) of data to be stored. The server sides are also enabled to access these storages without any restriction via the channel switches.

In addition, since the main site is connected with the remote site via a WAN, data can be shared between the sites, and a great amount of data can be shared worldwide. In addition, if a copy of data at the main and remote sites is retained at each other site, even when either site fails due to a disaster, etc., jobs can continue to run using the data at the other site. In this case, storages for backup data at the remote site are not limited to the same type of storage as at the main site, for example, not limited to copying from a RAID on the main side to a RAID on the remote side, and

hence cost reduction and simplified management may be achieved by copying from a RAID on the main side to a DVD-RAM or tape library, etc., on the remote side. In this case, the operation and management software on a terminal for managing a SAN manages the copy source, copy destination, etc., of these data.

In addition, in a prior art shown in Fig. 2, clients are connected with an application-specific server, for example, a main frame, a UNIX server, and a PC server, individually through communication lines such as a LAN, and individual servers are also connected via a LAN. Storages are connected with their respective servers. Therefore, data stored in the storages could be accessed only through their respective servers.

On the other hand, in the preferred embodiment of the present invention, data stored in storages connected with individual servers are managed in an integrated manner via a SAN. Firstly individuals of multiple servers are connected to various storages (such as a RAID disk drive, a tape library, and a DVD-RAM library/library array) via fiber channel switches (FC-Switches) of which the SAN is comprised. Thereby, data stored in individual storages are enabled to be accessed directly from individual servers without passing a LAN. For example, access to a great amount of data, etc., is simplified. In addition, since storages for data are consolidated into an integrated storage system, management of data and equipment is simplified.

In addition, in order to make backup and remote copies, etc., of data against a disaster, individual storages corresponding to each server must be installed and the data must be copied via a LAN according to a prior art, however, in the preferred embodiment of the present invention, an integrated storage system consisting

of a SAN and various storages is introduced, and hence the integrated storage system enables to back up data, and furthermore remotely and more efficiently.

As a computer system to which a SAN is applied is outlined above, the computer system must be an information system that is intended primarily for making any information about the data to be handled available at any time, for anyone, and from anywhere.

The integrated storage system relating to a preferred embodiment of the present invention, as disclosed in Fig. 3, firstly has as one of the basic functions the data protection that provides the backup as a measure against disk drive failures and the disaster recovery as a measure against a disaster such as an earthquake and fire, secondly has as one of the basic functions the data exchange and sharing among main frames, UNIX servers, and PC servers and the data sharing in which many types and forms of information such as a database (DB), documents, drawings, multi-media contents are handled, and lastly has as one of the basic functions the storage management (storage resource management) that provides unified management of storages that each server operated and managed separately, and the environment set-up and storage operation/management by standardized operations.

Concretely described below are details of individual basic functions according to the present invention. These functions are realized by installing a program (software), which describes these functions, and necessary data in memory of devices such as a storage, a switch, a server (computer), and a management unit (realized by a computer, etc.), and executing the program on a central processing unit (CPU) in theses devices individually. In addition, a data center in which a SAN-applied computer system consisting of a system group of a large capacity of storages

and various servers is connected to the Internet and is equipped with data storage service functions, namely Internet data center (abbreviated to "iDC"), is constructed, and an inventive device relating to a method for processing a mass of data at that iDC is one of features of the present invention.

First the data protection is described. Functions of the data protection are intended for backup of DBs during online operation, reduction in the management cost by sharing storage resources, improvement in system availability by means of disaster recovery, etc., and assurance of data security, and thereby, enable to back up data without stopping a job (non-disruptive backup) for 24-hour-per-day, 365-day-per-year operation that is expected to increase in the years ahead, enable to share a tape library at the time of backup (tape unit-shared backup), resulting in reduction in the cost as well, and further enable to restore the system rapidly in the event of a disaster by ensuring data security in copying remotely at long distance (remote copying). To put it concretely, the details of the data protection are three techniques of the non-disruptive backup, the tape unit-shared backup, and the asynchronous remote copying as described above.

Firstly functions or actions of the non-disruptive backup enable applications to run even during backup operation by the backup using a replica of data, and prevent application servers from being affected by using servers for backup only.

Fig. 4, Fig. 5a, and Fig. 5b illustrate a configuration for, and a function of the non-disruptive backup in detail. An outline of this function is to back up DBs without affecting online jobs via a SAN without passing a LAN by collaboration between internal functions in storages and database management system (DBMS) in

application servers.

Fig. 4 illustrates a series of a flow of the non-disruptive backup. First, by using said internal functions in storages, copying from the volumes to be backed up (primary volumes) to the secondary volumes with a capacity equal to or larger than that of the primary volume in a storage unit is executed to make a copy of the primary volumes. Next, during execution of applications, the status of the database management system (DBMS) in an application server is changed to a backup-allowable state to prevent online jobs from being affected, and then the backup server makes a backup copy of data in the secondary volumes to tape units.

Fig. 5a and Fig. 5b illustrate an outline of the processing by the volume copy function that is an internal function of a storage unit, in a process of the non-disruptive backup illustrated in Fig. 4. According to a prior backup technique not shown in the figure, originally, after stopping the jobs which a server performs to a database (DB), a backup copy of the DB is made to other storages, and after the relevant backup processing is complete, said online jobs to the DB is restarted. According to the prior art, online jobs to a DB must be in stop during backup operation.

In contrast to this, in one example of preferred embodiment of the present invention as illustrated in Fig. 5a, a replica for backup, namely Logical Volume B (Logical VOLB), is secured in a storages and a copy is made in advance. When backing up data in Logical Volume A (Logical VOLA), the data in Logical VOLA is copied to Logical VOLB in advance too. To put it concretely, if Logical VOLA is a backup target, two logical volumes of Logical VOLA and Logical VOLB are prepared

in advance and duplication is directed.

While data in Logical VOLA is being copied to Logical VOLB sequentially in the storage unit, when data is written to the storage unit from an online job (JOBA in the figure) concurrently with the copying, the duplicated writing of the data from the job is automatically performed on both Logical VOLA and Logical VOLB in the storage unit. After completion of copying sequentially from Logical VOLA to Logical VOLB, if data is written from JOBA, duplicated writing is also performed to keep individual data of Logical VOLA and Logical VOLB identical.

When performing backup, the backup server instructs the storage unit to perform pair split by using a means for controlling disk drives. After the split instruction, although data is written from JOBA, the storage unit writes the data to Logical VOLA only, and not to Logical VOLB. Thereby, data present in Logical VOLA when the split instruction is given is left in Logical VOLB as it is. After the split instruction, the backup software on the backup server reads data from the secondary volume, Logical VOLB, and makes a backup copy of the data to a backup device such as a tape unit.

However, for the volume duplication scheme illustrated in Fig. 5a, a duplicated volume must be prepared before a time when backup is performed. Therefore, in order to perform backup, volume duplication must be started further the duplication time before a backup time by taking into consideration the time taken to duplicate a volume. A function of a storage unit illustrated in Fig. 5b solves this problem.

In the case of Fig. 5b, Logical VOLB to which a copy of Logical VOLA is

made must be prepared in the same way as for Fig. 5a. Before starting backup, the backup server instructs the storage unit to perform pair split by using a means for controlling disk drives in the same way as for the case of Fig. 5a. However, at this time, data in Logical VOLA does not need to have been copied to Logical VOLB. After the split instruction, the backup software on the backup server starts reading data from the secondary volume, Logical VOLB. While data in Logical VOLA is being copied to Logical VOLB sequentially in the storage unit, if there in no data present in Logical VOLB when the backup server attempts to read data from the secondary volume, Logical VOLB, the disk drive reads out data from Logical VOLA and hands the data over to the backup server, or copies data from Logical VOLA to Logical VOLB once and then hands the data over to the backup server. As a result of this processing, although there is no data present in Logical VOLB at the time of splitting, it appears from view of the backup server that a copy of data in Logical VOLA is present in Logical VOLB.

However, data may be written from the application server into a certain area of Logical VOLA during the backup processing. Since data in Logical VOLA is being copied to Logical VOLB sequentially in the storage unit, if the data from the application server is written into Logical VOLB by the processing of copying, data after the split is also written into Logical VOLB. To prevent this, the storage unit reads Logical VOLA's data currently present in the area for which a write demand is made and writes the data out into Logical VOLB. After that, the storage unit writes into Logical VOLA the data which the application server demanded to write. As a result of this processing, data present in Logical VOLA only at the time of the split instruction is copied to Logical VOLB. With this method, data in the primary volume

(Logical VOLA) does not need to have been copied to the secondary volume (Logical VOLB) when the backup processing starts, that is, system operation in which a copy of volumes must be prepared in advance is not required, resulting in improvement of system operational ability.

Fig. 7 illustrates an example of installing a system constructed for the non-disruptive backup illustrated in Figs. 4, 5a, and 5b. The application server is equipped with DBMS and a means for controlling disk drives, and the backup server is equipped with backup software and a means for controlling disk drives. As an advance preparation, the means for controlling disk drives is installed, its configuration is set up, and operation of the means for controlling disk drives is checked. After that, when constructing an non-disruptive backup system, first a DBMS script (Logging in, Setting the backup mode, Terminating the backup mode, and Logging out) is created, a script (Pair split, Pair event wait, and Resynchronization) of the means for controlling disk drives in the application server is created, collaborated operation with the backup software is checked, and parameters for allocation of logical unit and the means for controlling disk drives are set.

In addition, in the case of another example of non-disruptive backup configurations illustrated in Fig. 6, the primary and secondary volumes created with the mirroring software are mirror split according to an instruction from the collaborating tool in the application server, and while backup is performed by using one volume (secondary volume), jobs are enabled to continue by using the other volume (primary volume). Then, after the backup terminates, resynchronization is performed. To put it concretely, the duplicated writing to the primary and secondary volumes is performed with the mirroring software in the application server, accessing

a DB is stopped with the collaborating tool (software) in the application server, and accessing the DB is restarted after mirror split is directed. Next, the backup copying of data in the secondary volume is started to a backup device such as a tape unit connected with the backup server by use of the collaborating tool (software) in the backup server. After that, the collaborating tool in the application server that is notified of completion of the backup from the collaborating tool (software) in the backup server directs mirror resynchronization and performs duplicated writing again.

Next, Fig. 8 and Fig. 9 illustrate the details of a configuration and function of the tape unit-shared backup. This function outlined is intended for reduction in the management cost of data that are scattered among many servers, and reduction in the load of a LAN with the result that high-speed backup is achieved. Further, by enabling a tape library to be shared among many server sides, the expansive library can be made the effective use of (compared with the case where a backup tape unit is installed for each disk drive), and by sharing a single tape library among multiple servers, backup data can be output directly to a tape unit via a SAN without passing a LAN, resulting in achievement of high-speed backup.

The left one of Fig. 8 illustrates conventional tape unit backup. Backup data is copied from each disk drive of individual servers via a LAN, through the backup server, to a tape unit, and hence data passes a LAN every backup case, a load is put on the LAN. Further, a load is also put on the backup server every backup case.

In accordance with a preferred embodiment of the present invention, in the case of LAN-free backup illustrated in the middle one of Fig. 8, the backup processing can be speeded up by copying data from a disk drive to a tape unit via a SAN, and

backup is achieved by use of servers without passing a LAN. When performing backup, a single type of server can be used, and hence the load of servers is reduced. In accordance with another preferred embodiment of the present invention, since server-less backup illustrated in the right one of Fig. 8 enables to copy data directly from disk drives to a tape unit, the backup processing can be speeded up and the load of servers can be reduced as well. In accordance with the preferred embodiment of the present invention as illustrated in the right one of Fig. 8, disk drives must be equipped with a capability of writing into tape units, tape units must be equipped with a capability of reading data from disk drives, FC switches must be equipped with a capability of writing from disk drives into tape units, or FC-SCSI multiplexers (described later in the explanation of Fig. 9) must be equipped with a capability of writing from disk drives into tape units if tape units are connected to the FC-SCSI multiplexers.

Fig. 9 illustrates another example of configurations for tape unit-shared backup. The configuration shown in Fig. 9 corresponds to LAN-free backup shown in the middle one of Fig. 8. In this configuration example, two or more nodes share a tape library concurrently and individual servers back up. In accordance with Fig. 9, Server C is different in functions from Servers A and B, has a backup manager installed for managing all over the backup, in addition to a backup agent necessary to perform a backup operation practically, and is equipped with functions of assigning a backup drive, etc. Here, the backup drive, for example, has three drives and assigns Drive 1 to Server A. When a backup demand is made from Server A, the backup drive is controlled so that a tape cartridge for storing is loaded onto Drive A. In addition,

drives may be assigned to servers in such a way that the backup manager manages the condition of drive usage, selects unused drives, and assigns a proper drive of them. In the structure shown in Fig. 9, a set of an FC-SCSI multiplexer and a backup drive corresponds to a tape library shown in Fig. 8.

Concrete operation of the tape unit-shared backup shown in Fig. 9 is described below. First, the agent on Server A demands the backup manager to mount a tape cartridge. Next, the manager receiving the demand mounts a tape cartridge onto any drive of a tape library. Then, the managers goes on to inform the agent on Server A of completion of mounting and the name of the drive onto which a tape cartridge has been mounted. Then, the agent on Server A performs backup actually. To put it concretely, Server A reads data from a storage, and writes the data into the mounted tape cartridge through an FC switch and an FC-SCSI multiplexer. Following this, after completion of backing up, the agent on Server A demands the manager to demount the tape cartridge. The manager instructs to demount the tape cartridge, and all the processing terminates.

Next, the following describes a configuration for and a function of asynchronous remote copying in the disaster recovery as a measure of data protection. This is intended for assurance of data security by copying remotely at long distance, for quick restoration of a system in the event of a disaster such as an earthquake, for duplication of a database to a remote site without affecting the performance of the main site, and for continuation of a job at the remote site in the event of a disaster.

Fig. 10 illustrates a system configuration for asynchronous remote copying. A main site and a remote site are located away long enough from each other

not to suffer from a disaster at the same time in the event of it and are connected through communication lines. When information is updated at the main site and the updating is complete, completion of the update is reported to a server (without waiting for reflecting information on the remote site, that is, asynchronously). Next, updated data is copied sequentially at a proper timing from the main site to the remote site; however, if data is not transferred in the same order the data was updated at the main site, updated data is sorted by the time sequence in a system at the remote site and then the data is copied with the sequence of update guaranteed (for example, if update data of receipt and payment of money are stored in reverse order, this can cause to force improper dealings in processing of remains).

Next, the following describes a configuration for and a function of high-speed replication between servers in data sharing. As shown in Fig. 11, when loading data between a DB on a main frame (backbone database with high reliability ensured) and a DB on UNIX/NT servers (for example, a database for which easiness in data handling is considered more important than reliability of data when performing the statistical processing of data, and on which hence source data necessary for the statistical processing is loaded from the main frame DB), intermediate files as a file of the main frame DB are set up, and the data is moved from the backbone DB to the intermediate files once (because specifications of the data loader of a UNIX server are not defined so as to read data directly from the backbone DB). Since the data in the intermediate files is converted to such a level that the data loader of a UNIX server can read, a replication of data is made in the DB on the UNIX server through pipes to prepare a DB for the required processing. At this time, data replication from the backbone DB to the DB on the UNIX server is done without passing a LAN, and

hence high-speed replication between servers can be achieved. Here, intermediate files can be a virtual volume that is created temporarily on semiconductor memory, namely cache memory, on the outside of magnetic disk drives. With cache memory, data can be transferred at a higher speed.

Furthermore, in order that UNIX servers or PC servers can construct a data warehouse easily, by installing in the UNIX servers or their attached units the software which is capable of performing easily and quickly in GUI base a series of the processing from extracting data from a variety of source DBs such as backbone DB, through converting and consolidating data, up to loading data, the time taken to transfer data can be shortened when constructing a data warehouse.

Next, the following describes a configuration for and a function of integrated operation and management of systems including storages. For computer systems that are large in size and is required to run 24-hour-per-day continuously, system management, in particular, storage management is considered important.

As a typical function of storage management, listed is monitoring for device failures, in particular, what part fails in a device. In addition, required are system maintenance work such as backing up data at each site periodically against a system crash, system setting modification work when volumes are added, and further data handling such as moving data in some volumes to other volumes when the performance drops due to load congestion in a particular volume. At that time, monitoring the condition of the load is also important management work. In a conventional system, one maintenance terminal is installed for each storage unit, and individual storages must be managed from their respective terminals.

In a means of storage integrated operation and management relating to a preferred embodiment of the present invention, all storage units can be managed by a single terminal.

Fig. 12 illustrates an example of backup operation and failure monitoring in a large-scale office system. In ordinary office environment, there are data used commonly within each department and data used commonly by all departments. In this example, there exist multiple client computers and multiple server computers on floor A, floor B, and floor C individually, and a mail server and a World Wide Web (WWW) server which are used commonly as a enterprise general system by all departments are prepared to provide their services to each department.

For a small-size data so that it is used by each department, in many cases individual departments can make a copy of their respective data for backup, so a backup device such as a tape unit is installed in individual departments. In addition, multiple large-scale storages to store a large-size data and a backup device such as a tape library are installed at a computer center, and each device at the center, each system on individual floor, and an enterprise general system are connected mutually via a Storage Area Network.

A centralized monitoring console monitors all devices on individual floor, in the enterprise general system and at the computer center, and all device failure reports are collected to the centralized monitoring console. Service personnel can identify easily what device a failure occurs in by seeing the console. When data is destroyed due to failures, the data can be recovered (restored) from a backup device. This restore processing can be also initiated from the centralized monitoring console.

In addition, the centralized monitoring console has such a function that service personnel leave the terminal unattended in some cases, so in such a case a mail is sent to a cellular phone, etc., of the service personnel from the centralized monitoring console to notify them.

The centralized monitoring console also directs how to operate backup and manages the backup. The frequency of backing up and the requirement of a destination of backing up vary with the kind of data individually. For example, data almost unnecessary to back up (for example, data updated very rarely) and data accessed by only a particular department or person do not need to be backed up frequently. Or, even if attempting to make a backup copy of all data at the same time zone, there is a limit to the number of backup devices. The centralized monitoring console rearranges the frequency of backing up, the time zone, or the destination of the backing up in accordance with the data or volume depending on the need of users, and automatically performs the backup processing individually.

Fig. 14 illustrates a diagrammatic view of the processing of setting up volumes. In the case of a large-scale storage unit, multiple disk drives are grouped to one or multiple apparent logical devices (LDEVs). In addition, the storage unit has multiple ports to connect to hosts or fiber channel switches, and which ports are allowed to access to individual LDEVs can be set and changed for the storage unit. When a host references an LDEV, the LDEV is recognized uniquely with the port identifier and logical unit number (LUN) of the storage unit. Hereafter, this set of a port identifier and an LUN is called the host address. In the storage unit, this host address is assigned to individual LDEVs and is made open to hosts.

From the centralized monitoring console, a host address is assigned to LDEVs, and the type of hosts that can access individual LDEVs is set. Since all hosts are connected to all storages via a storage area network, there is the risk that a host which is not allowed normally to access a storage gains an invalid access to the storage, so the type of hosts that can access individual LDEVs can be registered in the storage to prevent invalid access.

Fig. 13 illustrates an example of monitoring the performance of storages. The centralized monitoring console can watch the condition of the load of each volume. To put it concretely, the load condition is the number of times per second I/O operations are received, the ratio of read and write operations, the cache hit rate, etc. Generally, a load is very seldom put on all volumes evenly, and volumes with an extremely high load put on them or volumes with nearly no load put on them may present. Since the condition in which an one-sided load is put on particular multiple volumes can be monitored on the centralized monitoring console all at once, when watching this condition, a load is reallocated in such a way that part of data on heavy-loaded volumes is moved to light-loaded volumes, thereby operation plan can be drawn up easily so as to prevent the performance of a overall system from being dropped.

In addition, Fig. 15 illustrates an example of a case where a storage unit has the functions of reallocating volumes. Some storage units have a small capacity but a comparatively high speed of volumes, and other storage units have a large capacity but a low performance of volumes. In such a situation, it is better to move data which has a low access frequency to a large capacity of volumes, and data which has a high access frequency to a high speed of volumes. In the disk drives involved in

this case, individual logical devices (LDEVs) can be moved to other areas.

In addition, reallocation of volumes is invisible from hosts both during movement of the logical devices and after movement of the logical devices, and volumes can be handled in the same as before movement. Disk drives obtain the usage rate of logical devices as statistical information, and send the information to a centralized monitoring console. The centralized monitoring console predicts how the usage rate of logical devices changes when a logical device is moved based on the information, and presents the prediction to service personnel. Service personnel can draw a reallocation plan more easily than in the case of the previous figure based on the prediction. In addition, from the centralized monitoring console, service personnel can instruct to move the logical devices actually or not, or set in advance detailed conditions under which, when individual volumes are set in a certain state, the volumes are automatically moved.

In addition, there is FC switch management as a part of integrated system operation and management, and the FC switch management enables to make various settings of FC switches and to manage the status of zoning, etc. To put it concretely, it includes management such as the displaying of a fabric topology, the setting of FC switches' zoning, and the setting/displaying of various parameters in FC switches, and these items can be watched on the centralized monitoring console. Fig. 16 illustrates an example of configurations of a fabric switch (FC) lying between servers and storages with the switch divided into three zonings.

Next, on the whole configuration of a computer system relating to a preferred embodiment of the present invention described above, the following

describes an concrete example of cases where a terminal in which the operation and management software illustrated in Fig. 1 has been installed, namely a management terminal, manages and controls the whole configuration of a computer system.

To back up (Fig. 4), which volume in a storage is to be backed up must be determined. Usually, a server manages data which an application stores in a storage in units of files. On the other hand, a storage manages data in units of volumes.

Therefore, when backup is started, if the SAN management unit (terminal shown in Fig. 1, in which operation and management software has been installed) is asked to back up a file by a server, the SAN management unit obtains information to identify a file, information about a backup device (address on a SAN, etc.), a backup time, etc., from servers. Further, the SAN management unit obtains information to identify a volume in which the relevant files have been stored from storages. Next, the SAN management unit instructs a storage in which the relevant files have been stored to create a replica (secondary volume) of a volume to be backed up using the obtained two kinds of information. To put it concretely, the SAN management unit instructs a storage which has a volume in which the relevant files have been stored to assign another volume (secondary volume) for creating a replica of the relevant volume (primary volume) and to create the replica. In assigning the secondary volume, considerations must be taken so that a volume of at least the same capacity as that of the primary volume must be assigned to the secondary volume, and the SAN management unit must grasp how large capacity and what configuration of volumes individual storages have. When the creating of the secondary volume terminates, the SAN management unit, receiving this termination report, instructs the storage to split a pair of volumes, and instructs the backup server to make a backup copy of data from

the secondary volume to a backup device while keeping the primary volume occupied in the normal processing from servers. The backup server reads data in the secondary volume via the SAN, and transfers the read data to the backup device. When the backup processing terminates, this is reported to the SAN management unit from the backup server, and then the SAN management unit reports termination of the backup to an application that asked to back up. Note that a time at which to split a pair of volumes is the backup time described above. In addition, a destination on the SAN to which to transfer backup data is said address of the backup device on the SAN. Here, while communication of control information between the SAN management unit and storages can be performed from the SAN management unit, through a LAN, a server, and a SAN, to a storage as illustrated in Fig. 1, the SAN management unit not shown in the figure and storages are connected directly via a LAN, said control information can be communicated through this connection.

In the above description, the SAN management unit plays the central role to control reception of a backup demand, creation and split of a replica, the backup processing, and reporting of backup termination, however, software in an application server and software in a backup server exchange control information directly via a LAN, and thereby can realize the backup system without making use of a SAN management unit (Fig. 6). In this case, compared with the case where a SAN management unit is used, individuals of software in the two servers must collaborate, however, the SAN management unit described above is not required, and hence this scheme is considered to be suitable for a comparatively small-scale system.

In the backup system described above, data is backed up by transferring it

to a backup device through a backup server, however, backup can be controlled so that data is transferred directly from the secondary volume in a storage to a backup device via a SAN (direct backup) without passing a backup server. In the case where a SAN management unit is used, this backup is achieved by instructing a storage to transfer data in the secondary volume to a backup device after the SAN management unit recognizes that a replica has been created and split. This instruction includes the address of the backup device on the SAN, etc.

In addition, in the backup system described above, applications play the primary role to specify the backup file and the volume, however, for files and volumes which are updated frequently and require backup every day or every several hours, the load of applications can be reduced by specifying periodical backup for the management unit and the backup software in advance.

Next, the following describes an example of functions of a SAN management unit in the tape unit-shared backup (Fig. 8). In the case of the LAN-free backup, data backup related to individual servers is almost the same in backup operation as the backup described above. Differences from the above are that since data associated with multiple servers must be backed up, conflict of the backup processing among these multiple servers must be arbitrated, and so functions of arbitrating this conflict are required from the SAN management unit. For example, the SAN management unit is required to have functions of preventing access congestion in a tape library by instructing multiple servers to back up according to the schedule made out in advance, etc.

The following describes an example of controlling the zoning function illustrated in Fig. 16 as an example of operations of a SAN management unit. In Fig.

16, cluster servers are connected to storages through a fabric switch. Here, the fabric switch is divided logically, that is, is treated as multiple switches. Therefore, if the storage side output destination of the switch in Zoning 1 and the storage side output destination of the switch in Zoning 2 or Zoning 3 have been separated, cluster servers belonging to the switch in Zoning 1 can not gain access to the switch in Zoning 2 or Zoning 3, and invalid access to the storage side output destination of the switch in Zoning 2 or Zoning 3 from cluster servers belonging to the switch in Zoning 1 can be prevented.

Such set-up of zonings in the switch is enabled by connecting a fabric switch and an SAN management unit not shown in the figure through a LAN, etc. not shown in the figure, and setting up said zonings in the fabric switch according to an instruction from the SAN management unit, etc. In the case where a SAN management unit is not used, zonings can be set up in the fabric switch by using a dedicated console, etc., however, control information for zoning must be set at the location of said dedicated console each time cluster servers and storages are added, changed, or detached, resulting in inefficient operation. By using a SAN management unit and setting up zonings from the SAN management unit through communication, the operability is improved.

A few examples of operation of an SAN management unit are described above, however, when providing various functions of the data processing, the SAN management unit basically obtains from servers and storages the information about files and volumes to be processed, a operation timing, a destination to which to move data, etc., and instructs the devices required based on these pieces of information to

process files and volumes (replica creation, data copying, split of replica, backup copying, remote copying, etc.,) according to the operation timing. Individual devices perform their processing according to instructions from the SAN management unit, and return the result of processing. On as needed base, they can make the SAN management unit return the result to the client that asked to process.

To put it in order, a preferred embodiment of the present invention is considered to be composed of the following steps: step 1; an SAN management unit (terminal in which operation and management software has been installed as shown in Fig. 1) accepts a request for processing data in an integrated storage system from applications which run on individual application servers (this step can be replaced with another step at which the SAN management unit creates a demand for data on its own accord according to a schedule made out separately in advance), step 2; obtains information (information to identify the data to be processed, a operation time, a destination to which to move data, etc.,) necessary for processing the relevant data, step 3; determines the order in which the SAN management unit starts various kinds of functional software (software to execute replica creation, data copying, separation of replica, backup copying, remote copying, etc.,) which reside on storages, network switches, and servers based on said obtained information and makes out a schedule such as a start timing at which to execute the functional software (this step is considered to be a step for collaborating individuals of the functional software), step 4; starts individuals of the functional software actually according to the schedule, step 5; obtains results of execution from the functional software on individual devices (this result at step 4 may affect the result at step 3, namely a schedule), step 6; reports a result at step 5 to an application that asked to process data. Note that this process is

divided to these steps for convenience, and two steps of them can be combined, or any step can be subdivided into several sub steps as a separate step.

As described above, since a SAN management unit has functions of collaborating multiple pieces of functional software and operate them, the SAN management unit can realize easily complex functions that individuals of the functional software cannot achieve and the SAN management unit enables the more accurate data processing in an integrated storage system. On the other hand, complex functions can be achieved by creating a single piece of large software without collaborating multiple pieces of functional software, however, this leads to a situation in which separate pieces of software must be developed for each kind of the data processing, resulting in an inflexible system.

Next, the following describes how storage systems and storage area network techniques are used in a large-scale computer system, using a concrete example. Fig. 17 illustrates an example of configurations of an Internet data center (abbreviated to "iDC"), which has been expanding in the number of systems recently. The Internet data center is entrusted with Internet service providers (ISPs) and WWW servers of individual enterprises (this system is called "housing"), and provides network management and server operation and management. Further, it also provides value-added services such as web design, construction of an electronic commerce (EC) system, and addition of high-degree security. The Internet data center provides solutions together that solve problems in enterprises, which want to do Internet business, such as shortage of system staffs and their skill, and preparation of server installation places and networks.

Since high-priced equipment such as a high-speed network line is shared in an Internet data center, there is a feature that an Internet data center, in provider's place, can provide services to many enterprises at a low cost. In addition, users and enterprises which utilize an Internet data center are released from burdensome work such as backup and maintenance and deal with a business at a lower cost than running a system alone. However, since iDC runs many Internet environments and many pieces of application software that individual enterprises use, high-speed Internet backbone lines and many high-performance servers must be installed. In addition, these facilities must have high reliability and high security. In these environments, high-speed and highly functional storage systems are indispensable.

The following describes an example of applying storage area network techniques to a large-scale system such as an Internet data center.

Fig. 18 illustrates a schematic configuration diagram of an Internet data center to which a large-scale storage area network (SAN) is applied. Multiple server computers exist at each enterprise, storages such as a disk drive and a tape unit are consolidated to a few units, one or two-three units, and servers and disk drives/tape units are connected mutually through fiber channel switches. Although individual storage units must be connected to individual server computers in an environment in which a SAN does not exist, storage units can be shared by all computers through a SAN, and hence can be consolidated and managed. In addition, when adding storage units, the storage units can be added while a host computer is in online (in operation), so the addition does not affect jobs.

In addition, from the point of view of backup, storage consolidation

through a SAN plays an effective role. Here, Fig. 19 illustrates a schematic configuration diagram of an example of non-disruptive backup under a SAN environment at an Internet data center. In this figure, individual server computers, storages, and backup libraries of multiple enterprises are connected mutually via a storage area network. A management host exists on the SAN to manage storage devices and to operate backup. Data in each server computer, for example, Web contents on a WWW server and data used by an application server, have been consolidated and stored in storages on the SAN.

The demands for backup is considered to be varied depending on the circumstances of each host computer. For example, there are cases where it is desirable that a backup copy of data is taken every day at a time when a load of access to a host computer drops, that is, during a time zone such as midnight for which the number of times access is made to disk drives decreases, or it is desirable that in the case of a host computer which is very busy on the processing of an update type of transactions, the host computer determines a backup start time optionally according to the time and circumstances, such as a time when a flow of transactions breaks. The management host accepts those demands from individual host computers and manages backup processing properly. In addition, since 24-hour-per-day continuous operation is important at an Internet data center, interruption of processing on the host computer must be avoided and non-disruptive backup is mandatory. Described below briefly is an example of backup processing.

For example, if individual server computers want to make a backup copy at some timing once a day, the management host makes out a schedule of the backup

beginning and ending for individual server computers. For example, a backup operation for a WWW server of Company A begins at midnight, a backup operation for an application server of Company B at one in the morning, a backup operation for an application server of Company A at half past one in the morning, a backup operation for a WWW server of Company B at three in the morning, and so on. Time taken to perform the backup processing depends on the amount of data that individual servers keep, etc., and hence the management host manages what amount of data individual server computers keep in storages, and calculates the time taken for backup based on the amount of data and makes out a schedule. In addition, if a tape library has multiple tape drives, multiple backup jobs can be executed concurrently.

Taking as an example a case where a backup operation for Company A begins at midnight, the following describes a flow of processing. When midnight comes, the management host creates a replica of data, present in disk drives, of a WWW server of Company A. For that, the management host finds out a free disk (logical volume) in a disk drive, assigns it to a volume for the replica of a WWW server of Company A, and instructs the disk drive to create the replica. A flow of the processing of creating a replica is that as illustrated in detail in Fig. 5a and Fig. 5b.

Following this, a tape cartridge is mounted onto a tape drive in a tape library. After that, the copying of backup data begins from the replica volume to the tape library. The server computer of Company A can perform the data backup processing, however, if the direct backup function by which data is transferred directly from the management host or a disk drive to a tape library is supported (all right if at least any of a disk drive, a tape library, and a FC switch supports), this function can

actually be used for backup processing.

In that case, while the server computer is not aware of whether the backup processing is performed or not, a backup copy of data is automatically made. When the backup processing is complete, the tape cartridge is demounted from the tape drive, the replica volume in the disk drive is placed out of use, the volume is set to a free volume again, and the next backup processing follows.

In this case, since the tape library is shared and connected mutually via the SAN, if the schedule of tape library utilization is managed properly by the role of the management host, etc., one tape library can cover all their backup volumes even for multiple host computers. In addition, it is sufficient to prepare a replica volume only at the time the backup processing is needed if the management host assigns volumes properly, a replica volume does not need to be always prepared in individual volumes, and hence the number of tape library units and the number of volumes, etc., can be reduced.

Next, though the merits of sharing of storage units through a SAN are large in cost reduction, on the other hand, there are considerations to be taken in an environment in which servers of multiple enterprises coexist. One of them is security. All server computers can gain access to all storage units on a SAN via the SAN, so a server of Company C can look at data of Company A on the same SAN. Next, described below are examples of means by which to solve these problems.

Fig. 20 illustrates an environment in which server computers and storages of multiple enterprises coexist on a SAN at an Internet data center. Under the environment in which storages are shared by Companies A, B, and C as illustrated in

the figure, first zonings of an FC switch are set so that server computers of individual enterprises can gain access to a particular path only to storage units. Next, LUs that server computers of individual enterprises use are assigned to individual paths in the disk drives. For example, if Company B uses two logical units of LU1 and LU2, LUs 1 and 2 are assigned to the middle path, and if Company C uses LU0, LU 0 is assigned to the right path.

Further, there are multiple LUs on the same path and the LUs are shared by multiple servers, however, individual servers do not want to share in some case. For example, Company B secures the path to access LU 1 and LU 2 in Fig. 20, however, there may be a requirement in which only some particular one of Company B's servers is permitted to gain access to LU1. In that case, access limitation is done by use of the LUN. The WWN of a particular server of Company B is registered in a disk drive, and it can be set so that only a server whose WWN has been registered can gain access to LU1.

These zonings, path assignment, and access limitation in units of LUs are set on the centralized monitoring console. The topology of an FC switch is checked on the monitoring console, zonings are set based on the topology, further as many LUs as necessary are mapped on individual paths, and LUs that individual companies can use are registered. Furthermore, for LUs to which mutual access is not permitted within the same path, the centralized monitoring console obtains the WWNs of host computers that are permitted to access, sets them in a disk drive, and limits access in units of LUs.

Next, described below is an example of applying a computer system which uses an integrated storage system consisting of a SAN and various storages. In recent

years, merge and consolidation of enterprises have increased. As a result, this gives rise to the need to integrate computer systems among enterprises.

Fig. 21 illustrates an example of a large-scale computer system in which computer systems of multiple enterprises are connected mutually. Host computers among enterprises are connected through the Internet, and mutual utilization of data is achieved. In addition, by introducing storage area networks, storages in individual enterprises are organized so that they are also connected through a public switched network or leased lines.

From the point of view of computer system operation, integration of data is important. Usually, application databases that are used by individual enterprises are different, only straightforward mutual connection among devices does not make direct mutual use of data available. Therefore, generally, individual data from multiple databases must be consolidated and integrated to construct a new database..

In Fig. 21, Enterprises A and B individually have a backbone database by which transaction processing such as account processing is performed, and an database of information system by which analysis processing is performed in offline using data in the backbone database. In this example, the data of the backbone databases of Enterprise A and Enterprise B are integrated to create a data mart for various jobs. In some case, a large-scale data warehouse is constructed once, and then a small-scale data mart for various applications may be created from the data warehouse individually. In the case where does not exist an environment in which storages are connected mutually via a storage area network, when integrating databases, data must be moved through a host computer and a network. Usually, many databases which

enterprises want to share have a large capacity, and hence it takes a large amount of time to transfer data.

In the example in Fig. 21, a replica of Enterprise B's data is created by using a remote copying function in storages. A replica volume is split once at a frequency of once a day or once a week, etc., and a replication server reads data in the split replica volume to create various data marts. Replication servers exist separately from various types of DBMS of information system which make use of data marts. Storages are combined mutually via a storage area network, and a replica of a database can be created without putting any load on a host by using the remote copying function in storages. In addition, replication servers that creates data marts, and DBMS of information system can be realized on separate host computers individually, and hence the processing of creating data marts does not affect jobs of a backbone DB and a DB of information system.

According to the present invention, an integrated storage system can be constructed by reinforcing collaboration of components or functions of a storage system in which a SAN is used, and all various functions illustrated in Fig. 3 can be achieved.

Further, by connecting an integrated storage system to the Internet and applying the system to an Internet data center that keeps a large capacity of data and achieves utilization of the data, Internet information services can be provided efficiently in the cost and both of quantity and quality, and timely.

[What is claimed is]

[Claim 1]

A computer system comprising multiple client computers, multiple various servers, multiple various storages which keep data, a local area network (LAN) which connects said computers and said servers, a storage area network (SAN) which lies between said servers and said storages wherein:

said SAN forms a switched circuit network which is capable of connecting any said servers and any said storages through fiber channel switches, and said computer system comprising:

a terminal which is equipped with operation and management software which performs storage management including management of logical volumes in said various storages, data arrangement, and error monitoring, management of setting up said FC switches, and backup operation for data in said storages.

[Claim 2]

The computer system as claimed in claim 1 wherein said SAN is connected to SANs in other computer systems via a wide area network (WAN).

[Claim 3]

The computer system as claimed in claim 1 wherein, when a backup copy of data in the primary volume in a storage is made to a backup device in a non-disruptive manner, the secondary volume corresponding to the primary volume is created by internal functions in the relevant storage, copies are made from said primary volume to said secondary volume, the made copies are transferred to said

backup device via said SAN without passing said LAN, and thereby backup is achieved.

[Claim 4]

A computer system comprising multiple client computers, multiple various servers, multiple various storages which keep data, a local area network (LAN) which connects said computers and said servers, a storage area network (SAN) which lies between said servers and said storages wherein:

said SAN forms a switched circuit network which is capable of connecting any said servers and any said storages through fiber channel switches, and

when a backup copy of data in the primary volume in a storage is made to a backup device in a non-disruptive manner, the relevant storage has functions of receiving a volume split instruction from servers, functions of assuming as if data in the primary volume were kept in the secondary volume as it is at the time of the relevant instruction, and functions of making a backup copy from said secondary volume to a backup device.

[Claim 5]

A method for managing a system consisting of servers, storages which store data of said servers, a network which connects said servers and said storages, and backup devices which are connected with said network and back up said data, and said method comprising:

step 1 to obtain the information to identify data to be backed up;

step 2 to obtain a specification of processing the data denoted by said

information;

step 3 to instruct said storage which keeps the data denoted by said information to execute said a specification of processing; and

step 4 to receive the result of processing the data denoted by said information from said storage.

[Claim 6]

The method for managing a system as claimed in claim 5 wherein, said specification of processing is to transfer said data from said storage to said backup device.

[Claim 7]

The method for managing a system as claimed in claim 5 wherein, said a specification of processing is to create a copy of the data denoted by said information, and to transfer said created copy data to said backup device.

[Claim 8]

The method for managing a system as claimed in claim 5, further comprising step 5 to obtain information about a timing at which said specification of processing is executed and step 6 to control execution timing at said step 3 according to said information about a timing.

[Claim 9]

The method for managing a system as claimed in claim 5 wherein, said

servers in said system are connected with the Internet, and said data is sent out to said Internet.

[Claim 10]

The method for managing a system as claimed in claim 6 wherein, said servers in said system are connected with the Internet, and said data is sent out to said Internet.

[Claim 11]

The method for managing a system as claimed in claim 7 wherein, said servers in said system are connected with the Internet, and said data is sent out to said Internet.

[Claim 12]

The method for managing a system as claimed in claim 8 wherein, said servers in said system are connected with the Internet, and said data is sent out to said Internet.

[Abstract]

In order to construct an integrated storage system by reinforcing collaboration of components or functions of a storage system in which a storage area network (SAN) is used, in a computer system comprising multiple client computers, multiple various servers, multiple various storages which keep data, a local area network (LAN) which connects the computers and the servers, a storage area networks (SAN) which lies between the servers and said storages, the SAN forms a switched circuit network which is capable of connecting any servers and any storages through fiber channel switches, and the computer system further comprises a terminal which is equipped with operation and management software which performs storage management including management of logical volumes in the various storages, data arrangement, and error monitoring, management of setting up said FC switches, and backup operation for data in said storages.

FIG.1

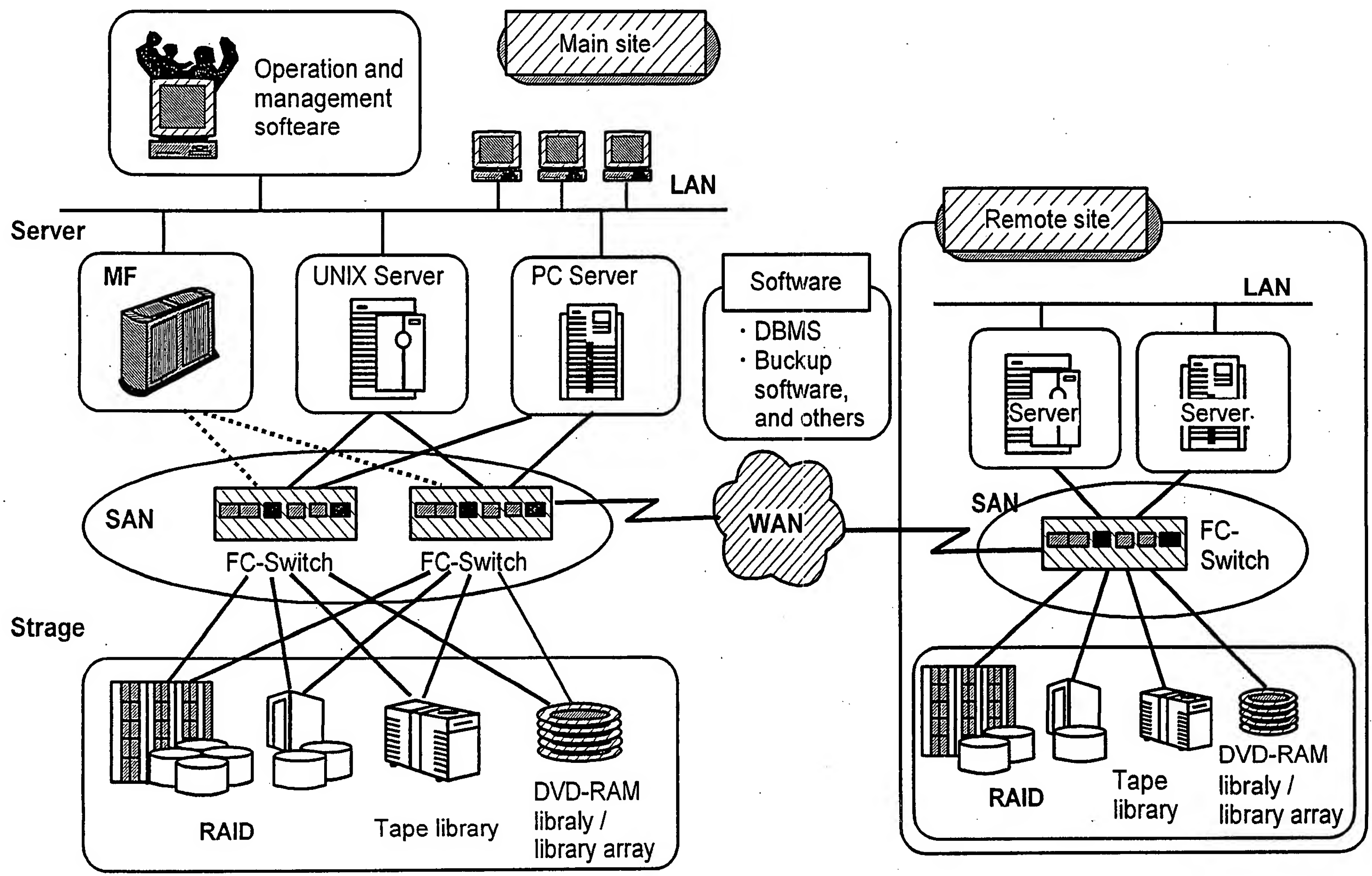
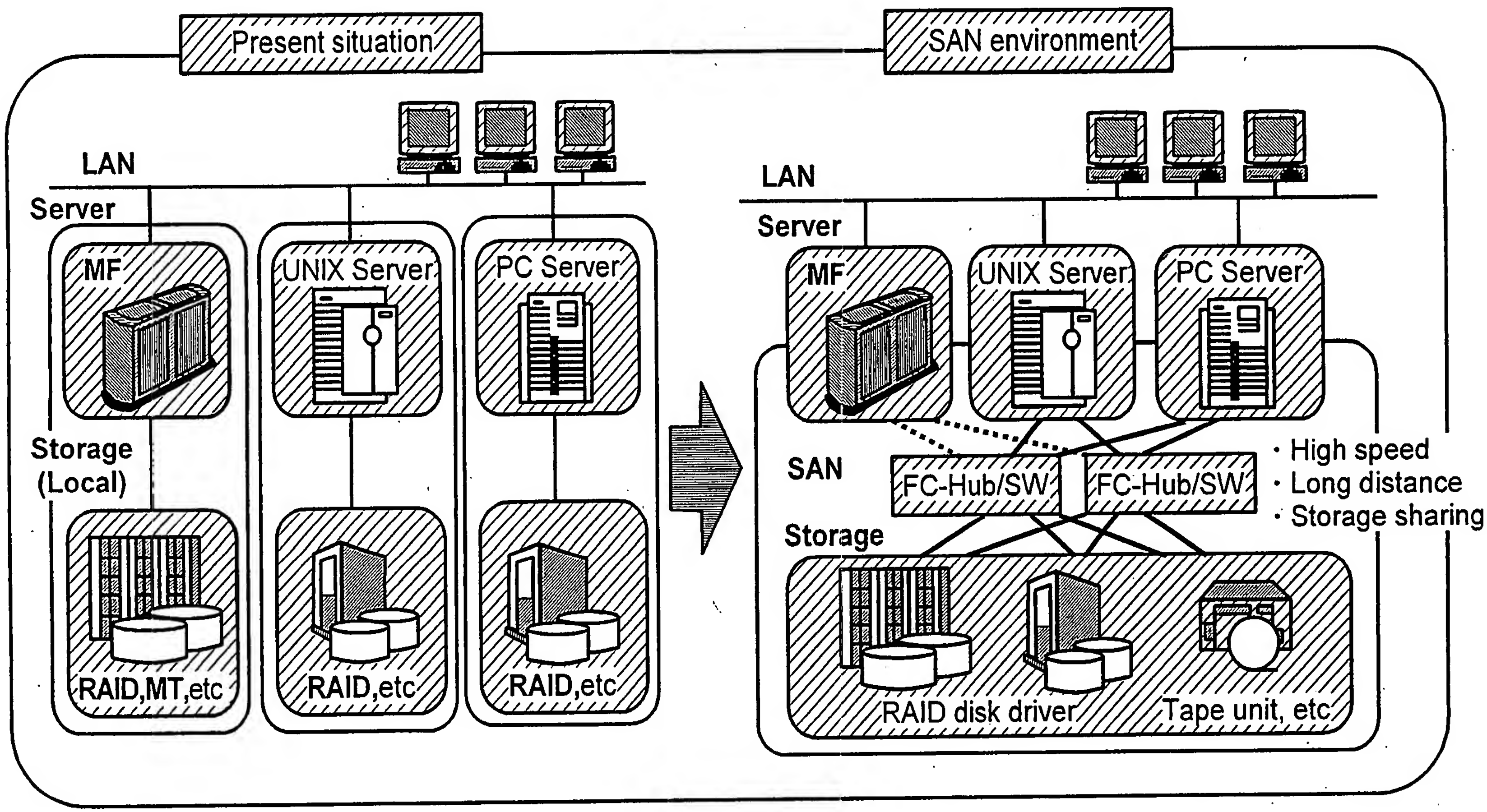

Operation and management softeare
Main site
LAN
Server
MF
UNIX Server
PC Server
Software
· DBMS
· Buckup software, and others
SAN
FC-Switch
FC-Switch
WAN
Strage
RAID
Tape library
DVD-RAM libraly / library array
Remote site
LAN
Server
Server
SAN
FC-Switch
RAID
Tape library
DVD-RAM libraly / library array

FIG.2



Present situation
LAN
Server
MF
UNIX Server
PC Server
Storage
(Local)
RAID,MT,etc
RAID,etc
RAID,etc
SAN environment
LAN
Server
MF
UNIX Server
PC Server
SAN
FC-Hub/SW
FC-Hub/SW
· High speed
· Long distance
· Storage sharing
Storage
RAID disk driver
Tape unit, etc

FIG.3

Primary Functions of Integrated Storage System

- Data protection (failure measures, disaster measures)
  - Backup (measures against failures in disk drives)
  - Disaster recovery (measures against disasters such as an earthquake and fire)
- Data Sharing (data sharing)
  - Data exchange and sharing among main frames, UNIX server, and NT servers
  - Processing many kinds of information (DB, documents, drawings, multimedia contents, etc.)
- Storage Management (storage resource management)
  - Unified management of storage resources that multiple departments manage and operate separately
  - Environment set-up and storage operation / management by standardized operations

FIG.4

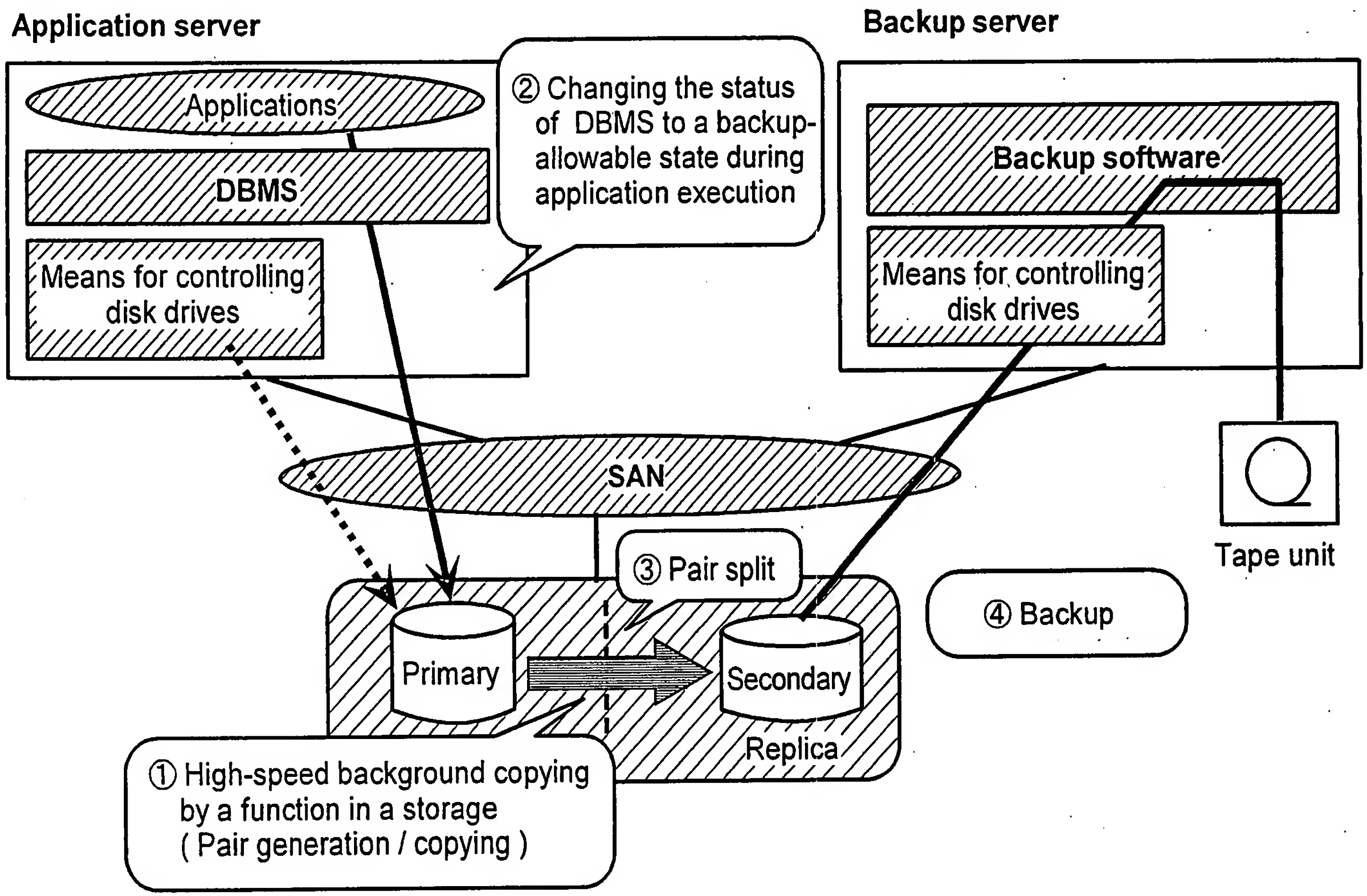
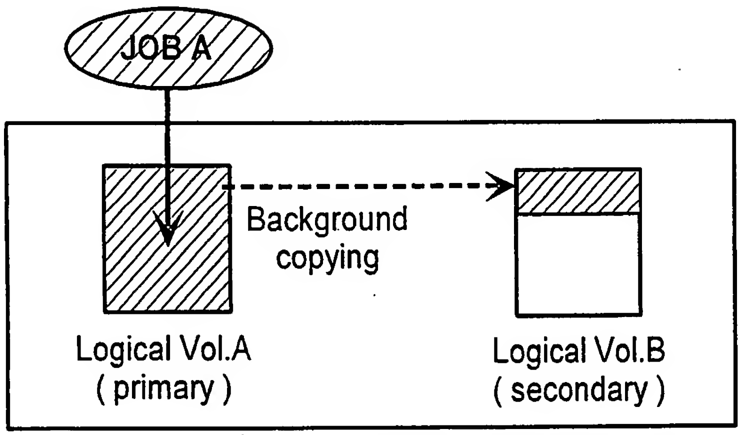
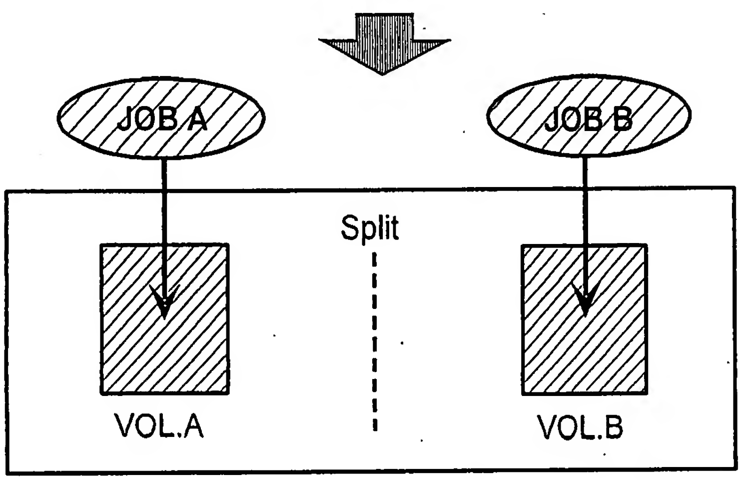

Application server
Applications
DBMS
Means for controlling disk drives
② Changing the status of DBMS to a backup-allowable state during application execution
Backup server
Backup software
Means for controlling disk drives
SAN
Tape unit
③ Pair split
④ Backup
Primary
Secondary
Replica
① High-speed background copying by a function in a storage (Pair generation / copying)

FIG.5a

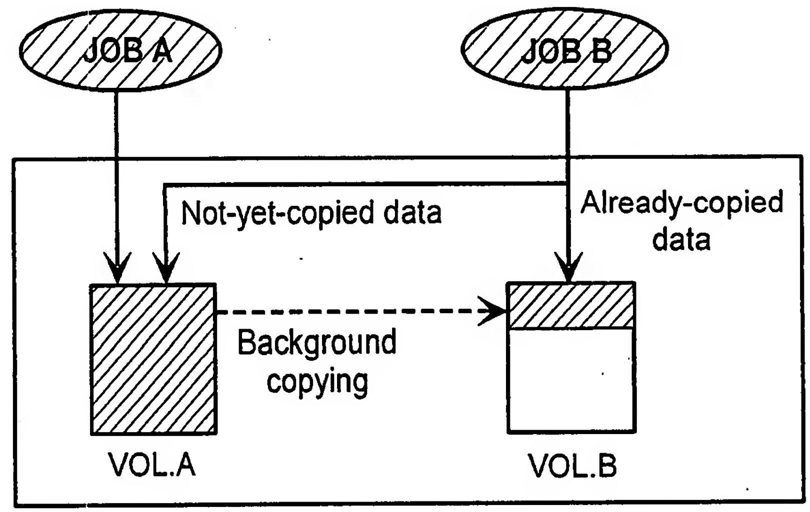


Makes a copy in the background in advance.



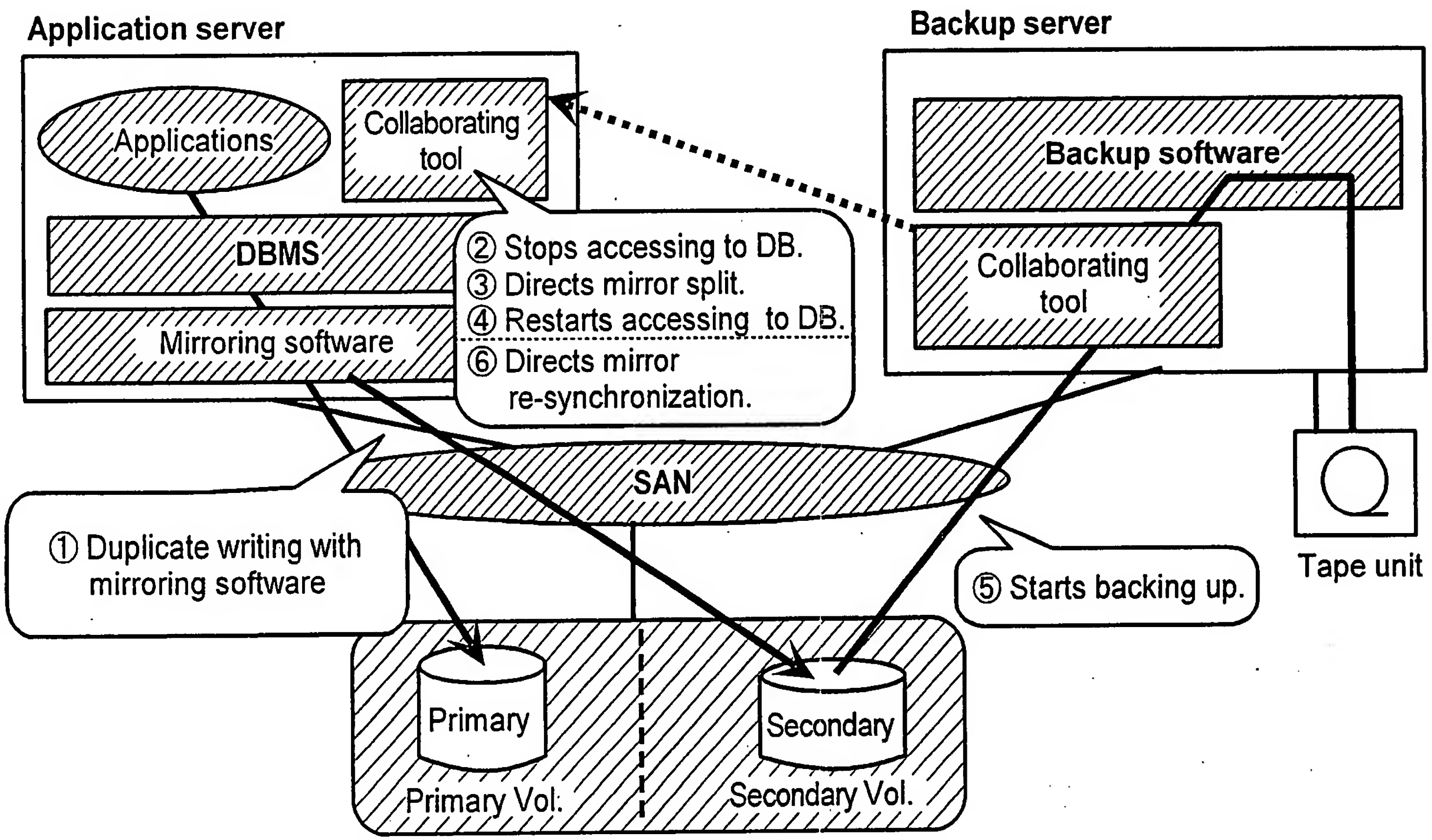
After copying is complete, the secondary volume is split and is available for another job to use.

FIG.5b



- Apparently, the secondary volume is split immediately and is available for another job to use.
- Actual data is copied from the primary volume in the background.

FIG.6



Application server
Applications
Collaborating tool
DBMS
Mirroring software
② Stops accessing to DB.
③ Directs mirror split.
④ Restarts accessing to DB.
⑥ Directs mirror re-synchronization.
Backup server
Backup software
Collaborating tool
Tape unit
SAN
① Duplicate writing with mirroring software
⑤ Starts backing up.
Primary
Primary Vol.
Secondary
Secondary Vol.

FIG.7

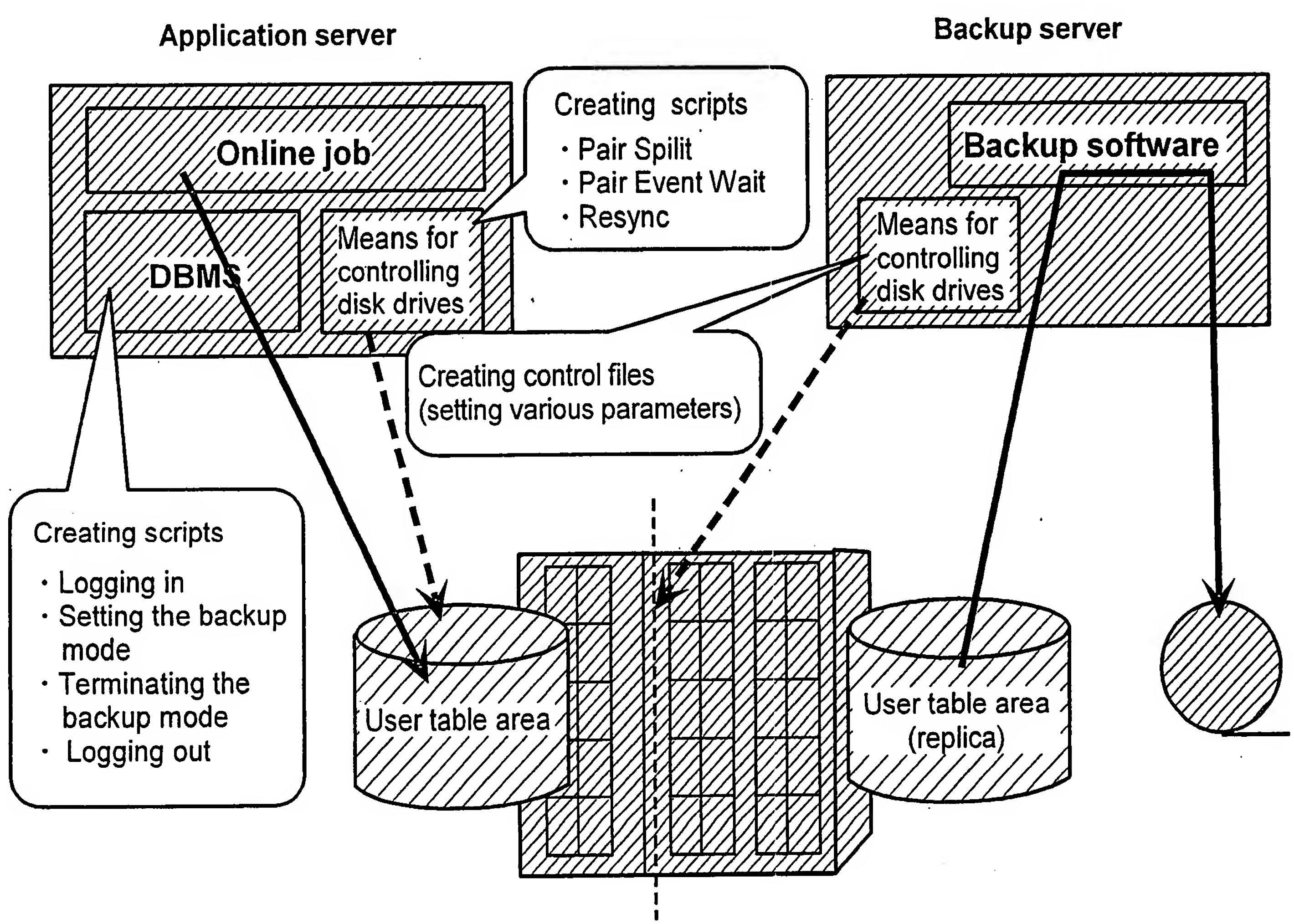

Application server
Backup server
Online job
DBMS
Means for controlling disk drives
Creating scripts
· Pair Spilit
· Pair Event Wait
· Resync
Backup software
Means for controlling disk drives
Creating control files (setting various parameters)
Creating scripts
· Logging in
· Setting the backup mode
· Terminating the backup mode
· Logging out
User table area
User table area (replica)

FIG.8

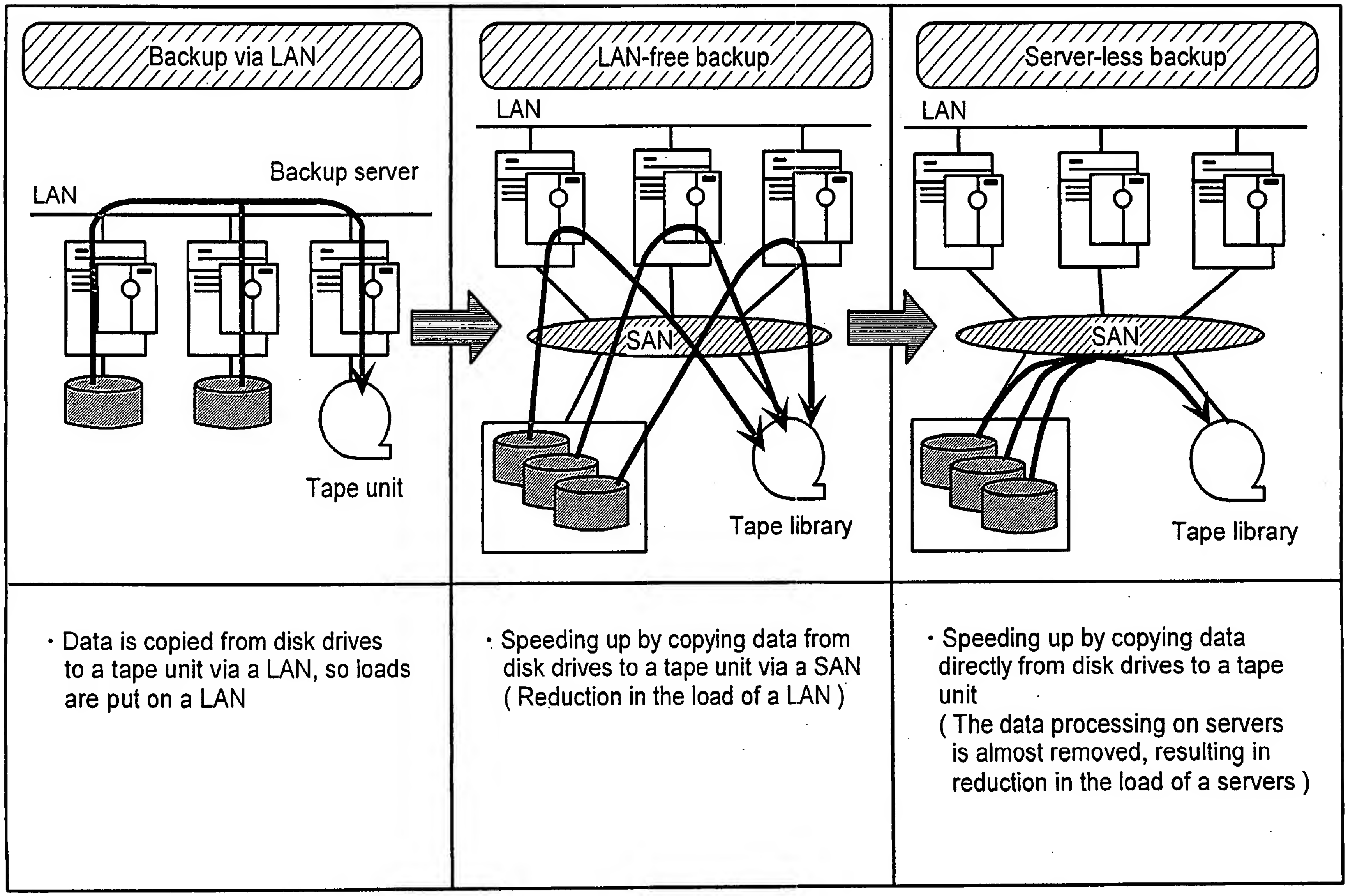

Backup via LAN
LAN
Backup server
Tape unit
LAN-free backup
LAN
SAN
Tape library
Server-less backup
LAN
SAN
Tape library
· Data is copied from disk drives to a tape unit via a LAN, so loads are put on a LAN
· Speeding up by copying data from disk drives to a tape unit via a SAN (Reduction in the load of a LAN)
· Speeding up by copying data directly from disk drives to a tape unit (The data processing on servers is almost removed, resulting in reduction in the load of a servers)

FIG.9

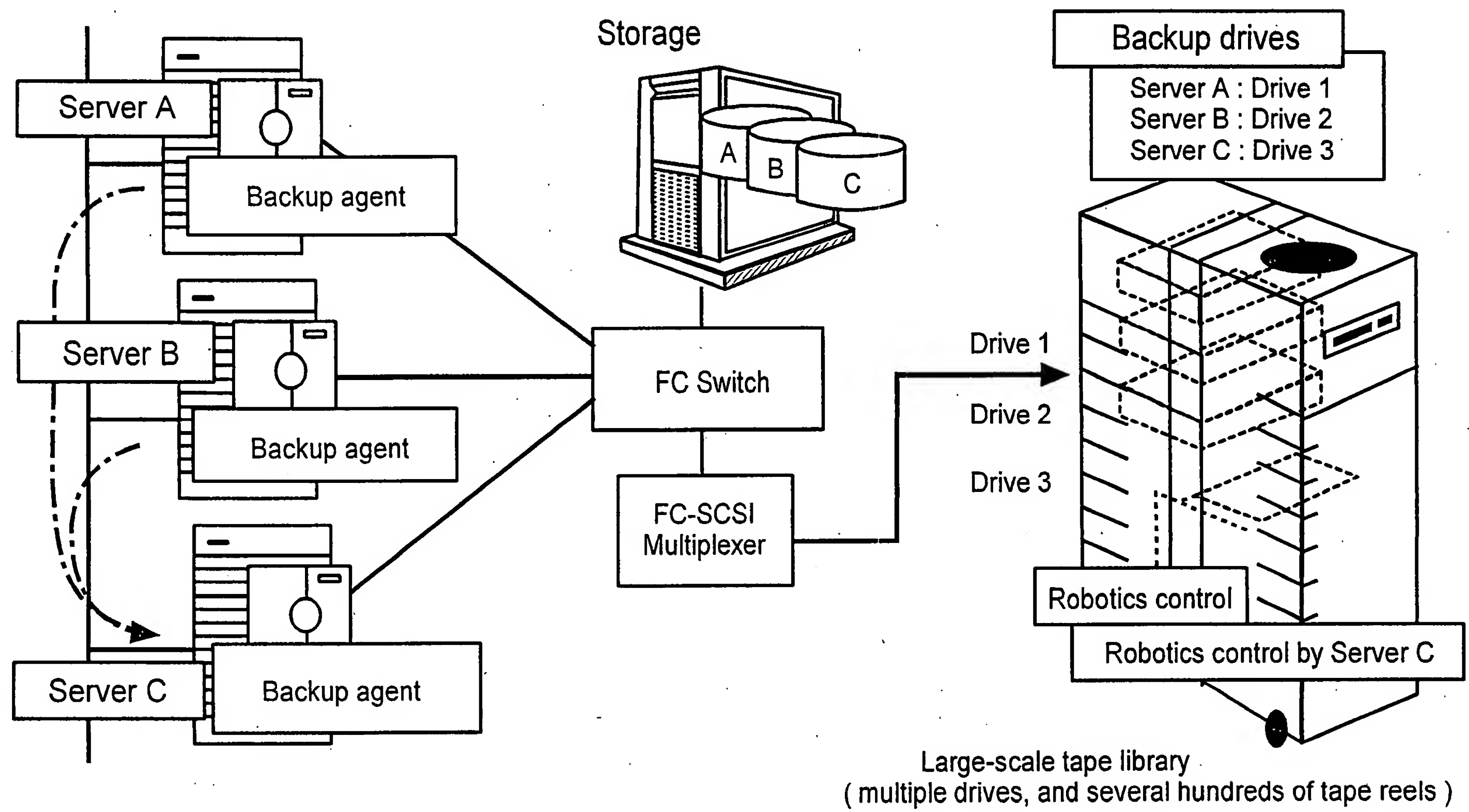
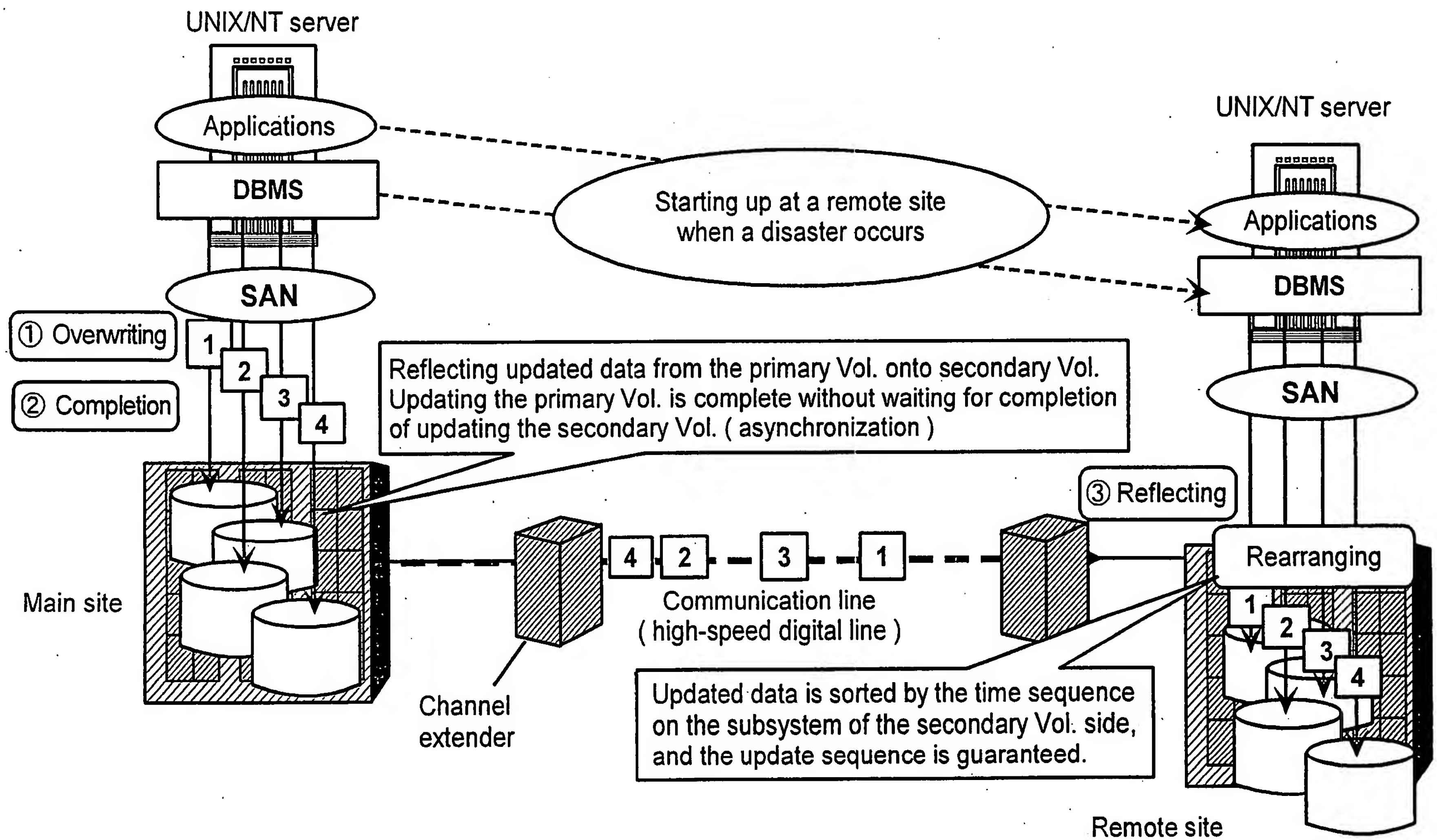

Storage
Backup drives
Server A : Drive 1
Server B : Drive 2
Server C : Drive 3
Server A
Backup agent
Server B
Backup agent
Server C
Backup agent
A
B
C
FC Switch
FC-SCSI
Multiplexer
Drive 1
Drive 2
Drive 3
Robotics control
Robotics control by Server C
Large-scale tape library
(multiple drives, and several hundreds of tape reels)

FIG.10



UNIX/NT server
Applications
DBMS
SAN
① Overwriting
② Completion
1
2
3
4
Main site
Starting up at a remote site
when a disaster occurs
Reflecting updated data from the primary Vol. onto secondary Vol.
Updating the primary Vol. is complete without waiting for completion
of updating the secondary Vol. (asynchronization)
Channel
extender
4
2
3
1
Communication line
(high-speed digital line)
③ Reflecting
Updated data is sorted by the time sequence
on the subsystem of the secondary Vol. side,
and the update sequence is guaranteed.
UNIX/NT server
Applications
DBMS
SAN
Rearranging
Remote site

FIG.11

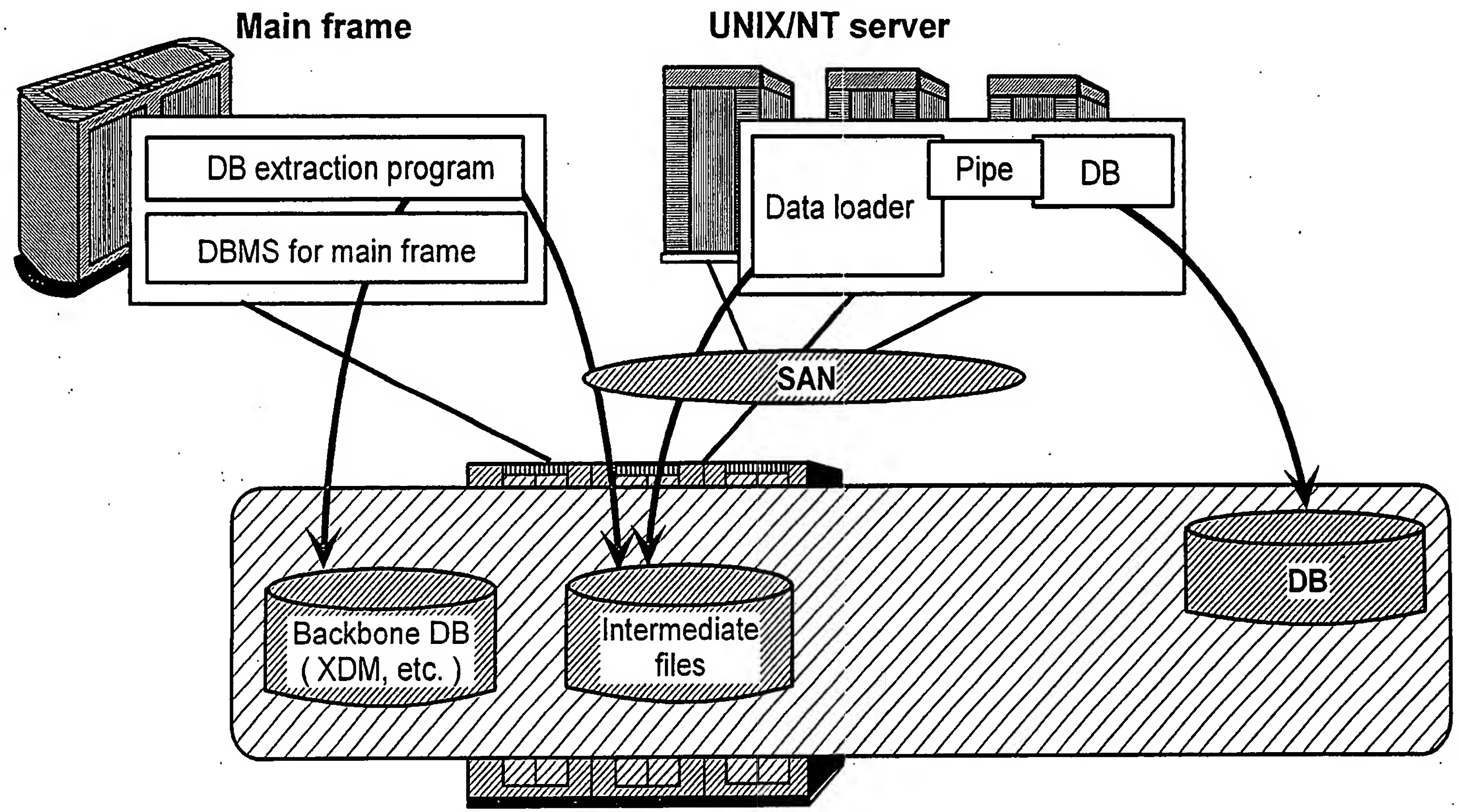

Main frame
DB extraction program
DBMS for main frame
UNIX/NT server
Data loader
Pipe
DB
SAN
Backbone DB
(XDM, etc.)
Intermediate
files
DB

FIG.12

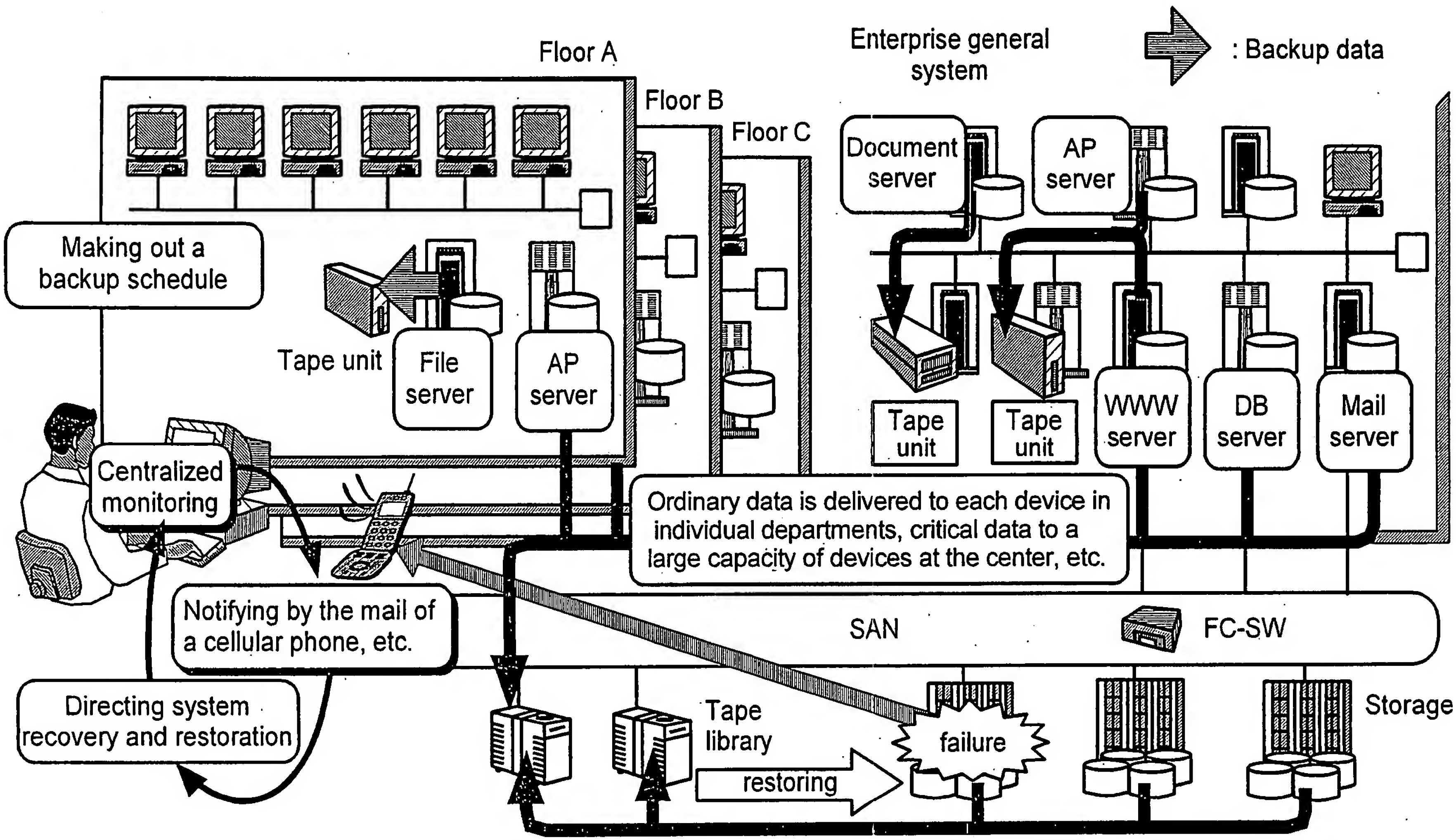

Floor A
Floor B
Floor C
Enterprise general system
: Backup data
Making out a backup schedule
Tape unit
File server
AP server
Document server
AP server
Tape unit
Tape unit
WWW server
DB server
Mail server
Centralized monitoring
Ordinary data is delivered to each device in individual departments, critical data to a large capacity of devices at the center, etc.
Notifying by the mail of a cellular phone, etc.
SAN
FC-SW
Directing system recovery and restoration
Tape library
restoring
failure
Storage

FIG.13

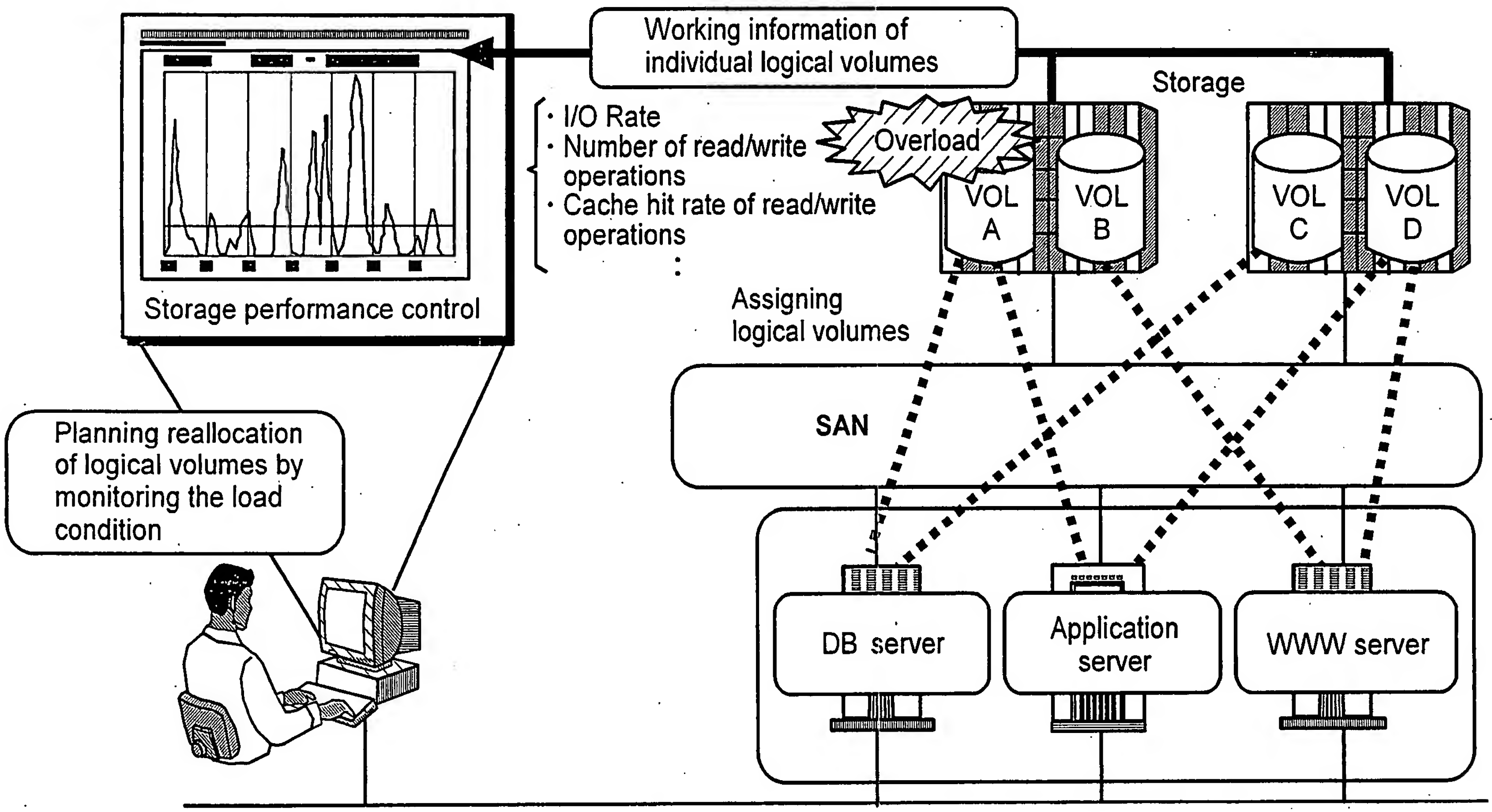
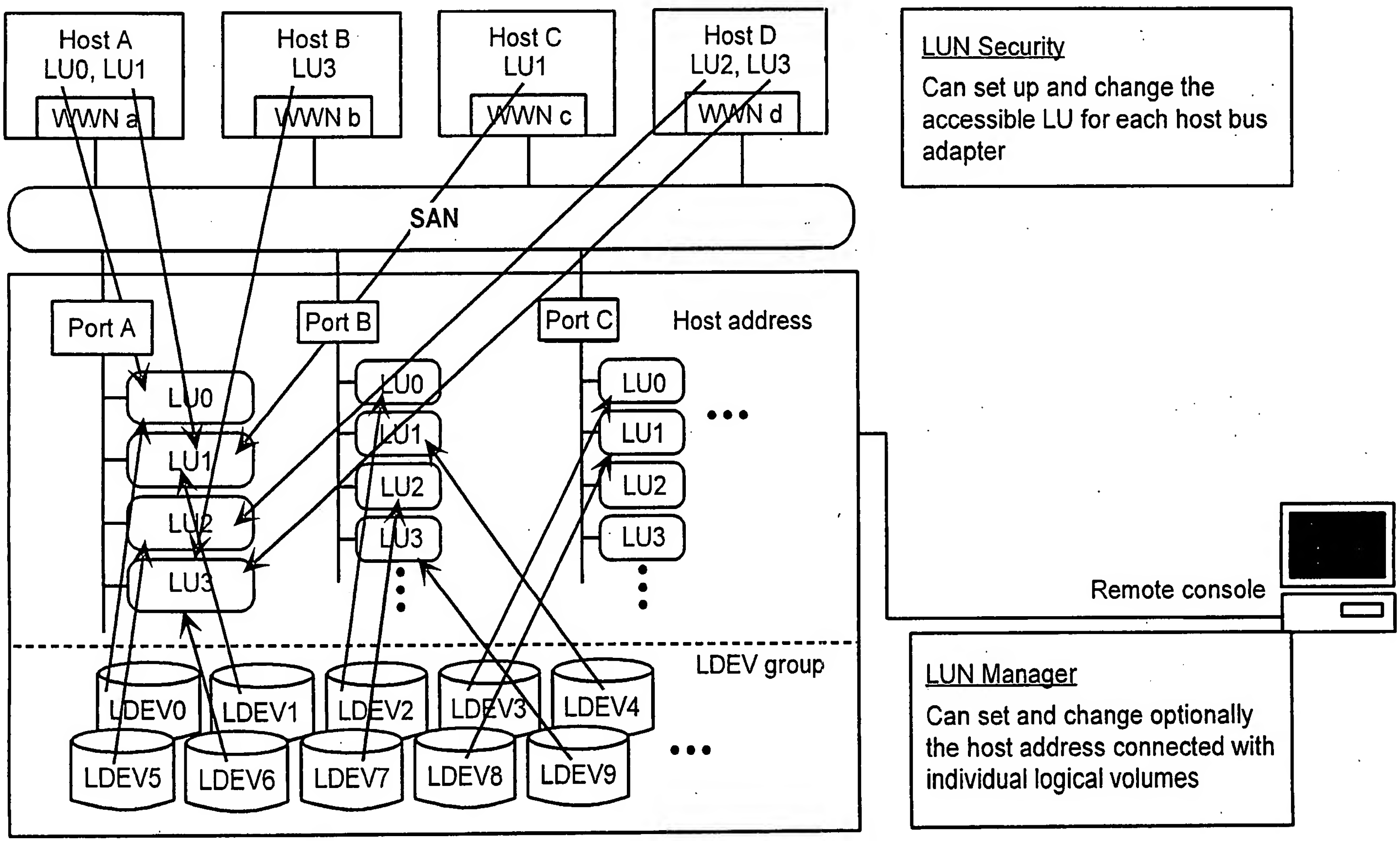
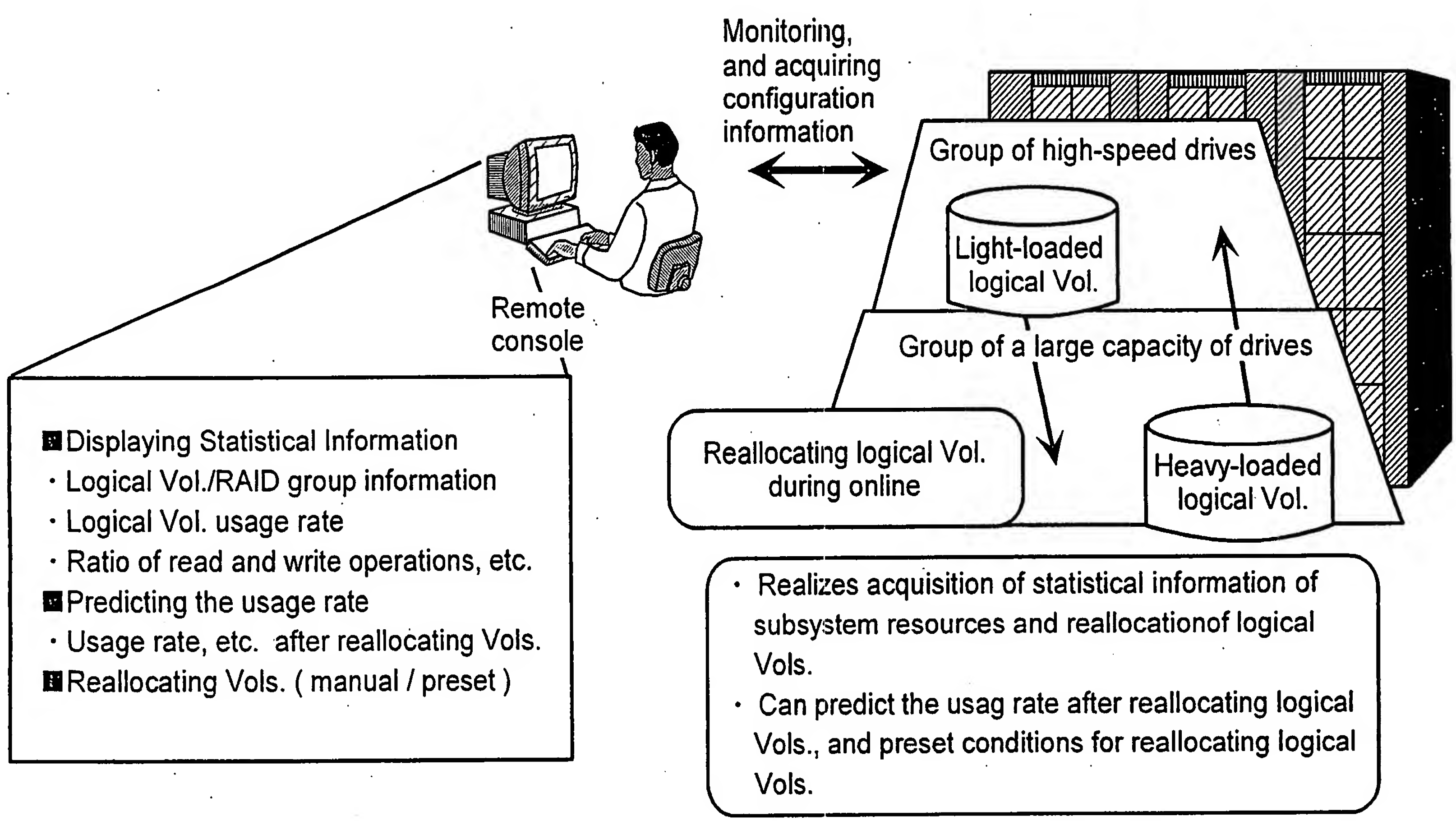

Working information of individual logical volumes
Storage
· I/O Rate
· Number of read/write operations
· Cache hit rate of read/write operations
:
Overload
VOL A
VOL B
VOL C
VOL D
Storage performance control
Assigning logical volumes
SAN
Planning reallocation of logical volumes by monitoring the load condition
DB server
Application server
WWW server

FIG.14



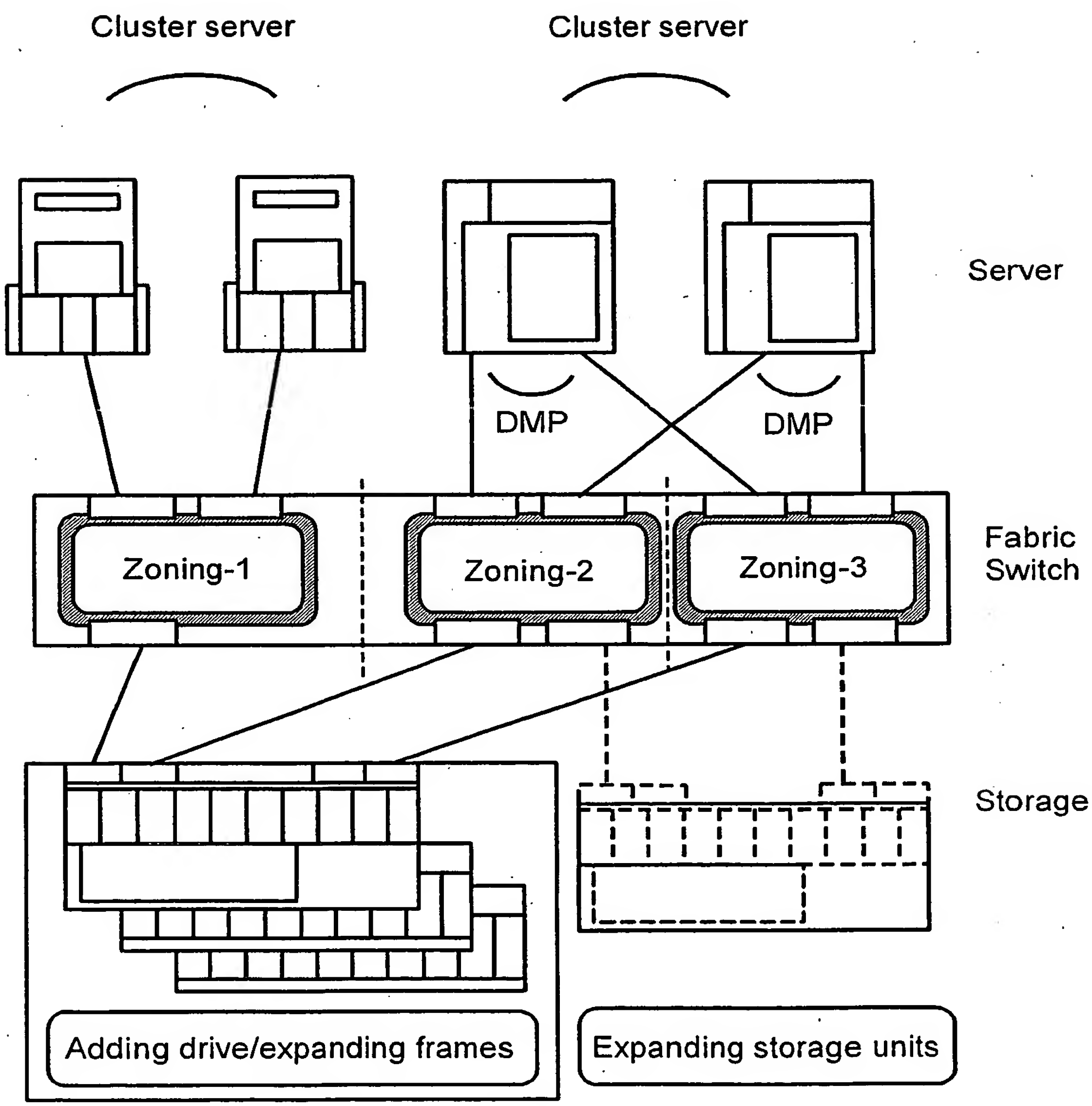
LUN : Logical Unit Number, LDEV : Logical device management logical volume

FIG.15



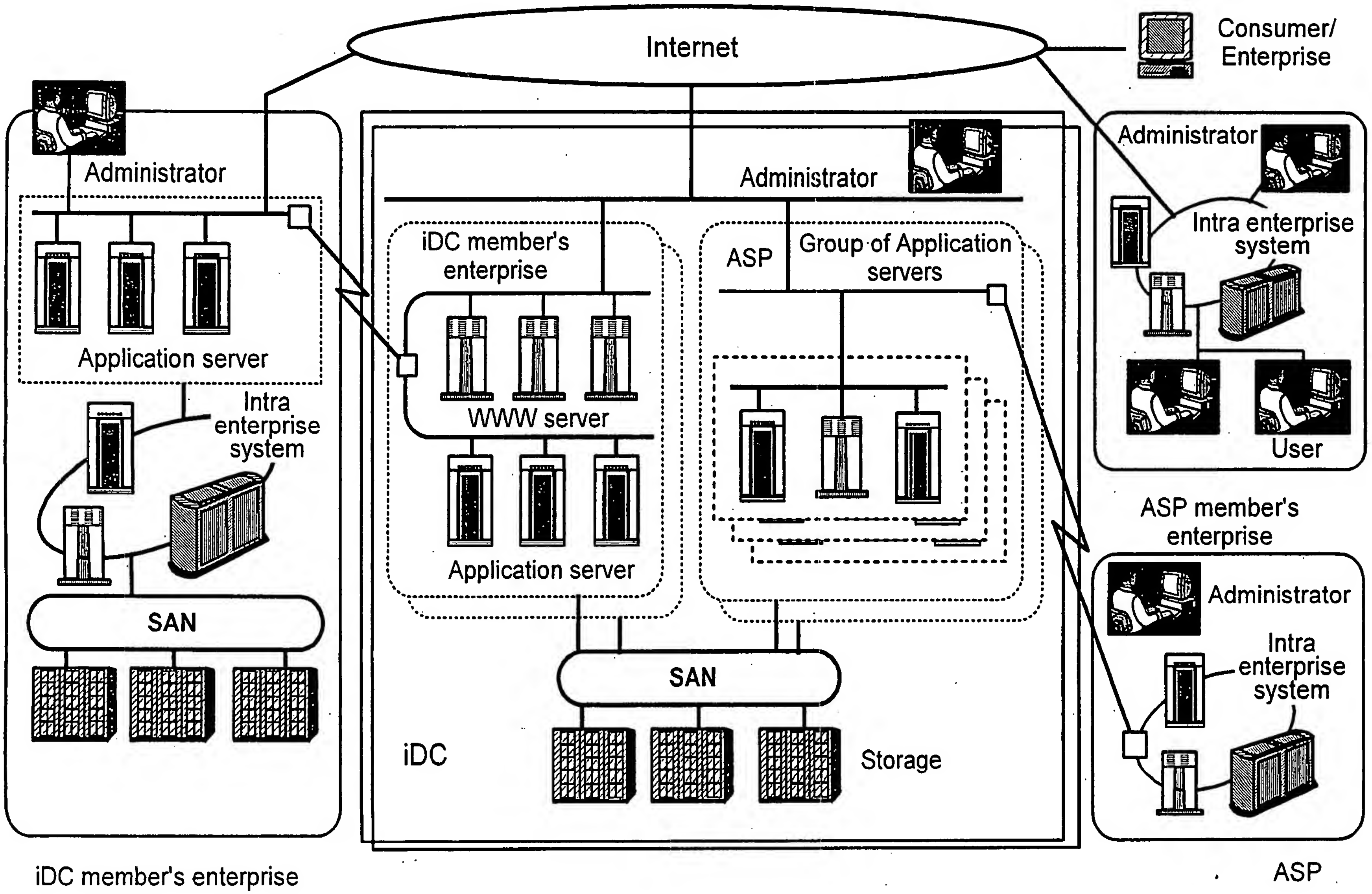
Monitoring,
and acquiring
configuration
information
Remote
console
Group of high-speed drives
Light-loaded
logical Vol.
Group of a large capacity of drives
Heavy-loaded
logical Vol.
Reallocating logical Vol.
during online
■ Displaying Statistical Information
· Logical Vol./RAID group information
· Logical Vol. usage rate
· Ratio of read and write operations, etc.
■ Predicting the usage rate
· Usage rate, etc. after reallocating Vols.
■ Reallocating Vols. (manual / preset)
· Realizes acquisition of statistical information of subsystem resources and reallocationof logical Vols.
· Can predict the usag rate after reallocating logical Vols., and preset conditions for reallocating logical Vols.

FIG.16



Cluster server
Cluster server
Server
DMP
DMP
Zoning-1
Zoning-2
Zoning-3
Fabric Switch
Storage
Adding drive/expanding frames
Expanding storage units

FIG.17



Internet
Consumer/ Enterprise
Administrator
Application server
Intra enterprise system
SAN
iDC member's enterprise
Administrator
iDC member's enterprise
WWW server
Application server
ASP
Group of Application servers
SAN
iDC
Storage
Administrator
Intra enterprise system
User
ASP member's enterprise
Administrator
Intra enterprise system
ASP

FIG.18

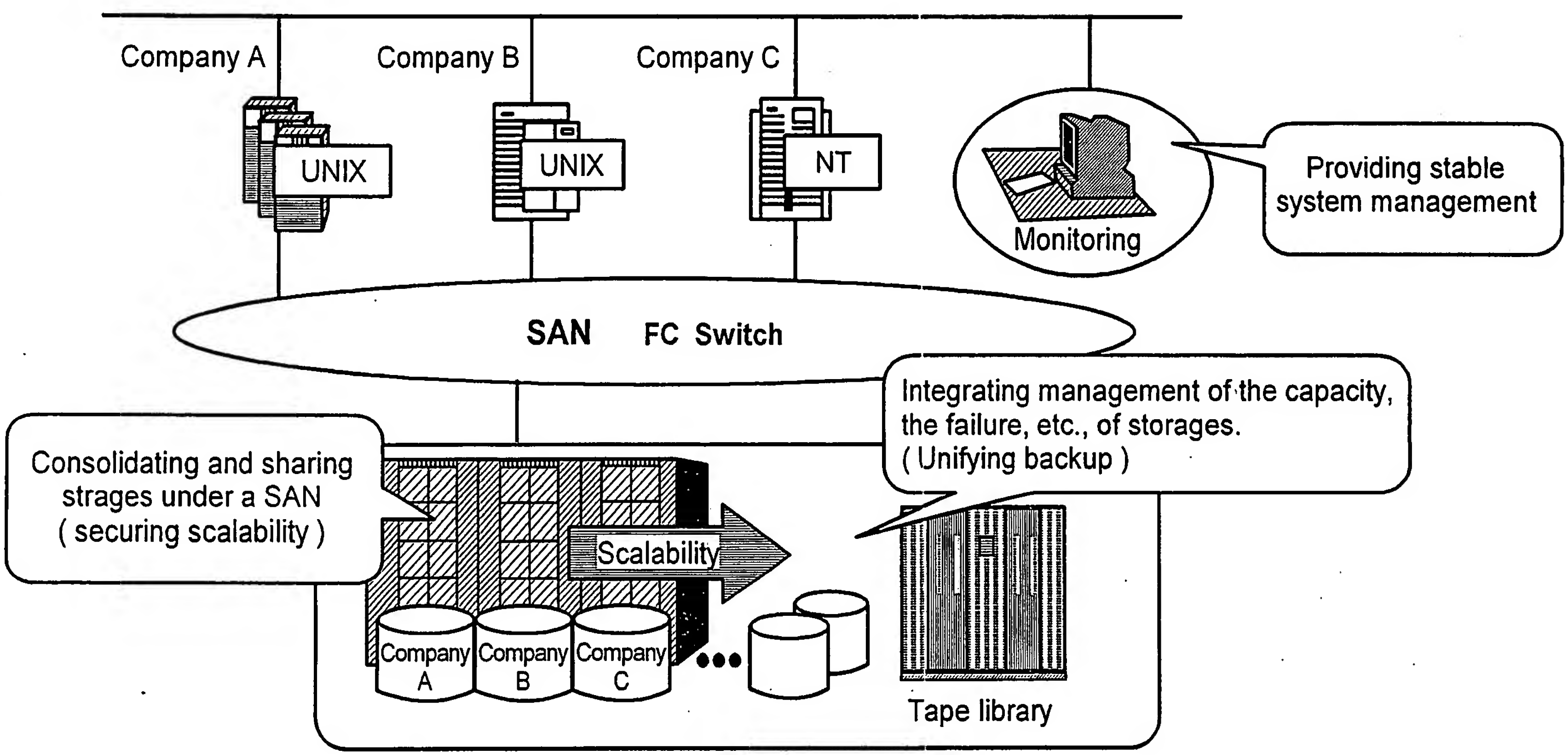

Company A
Company B
Company C
UNIX
UNIX
NT
Monitoring
Providing stable system management
SAN
FC Switch
Consolidating and sharing strages under a SAN (securing scalability)
Integrating management of the capacity, the failure, etc., of storages. (Unifying backup)
Scalability
Company A
Company B
Company C
Tape library

FIG.19

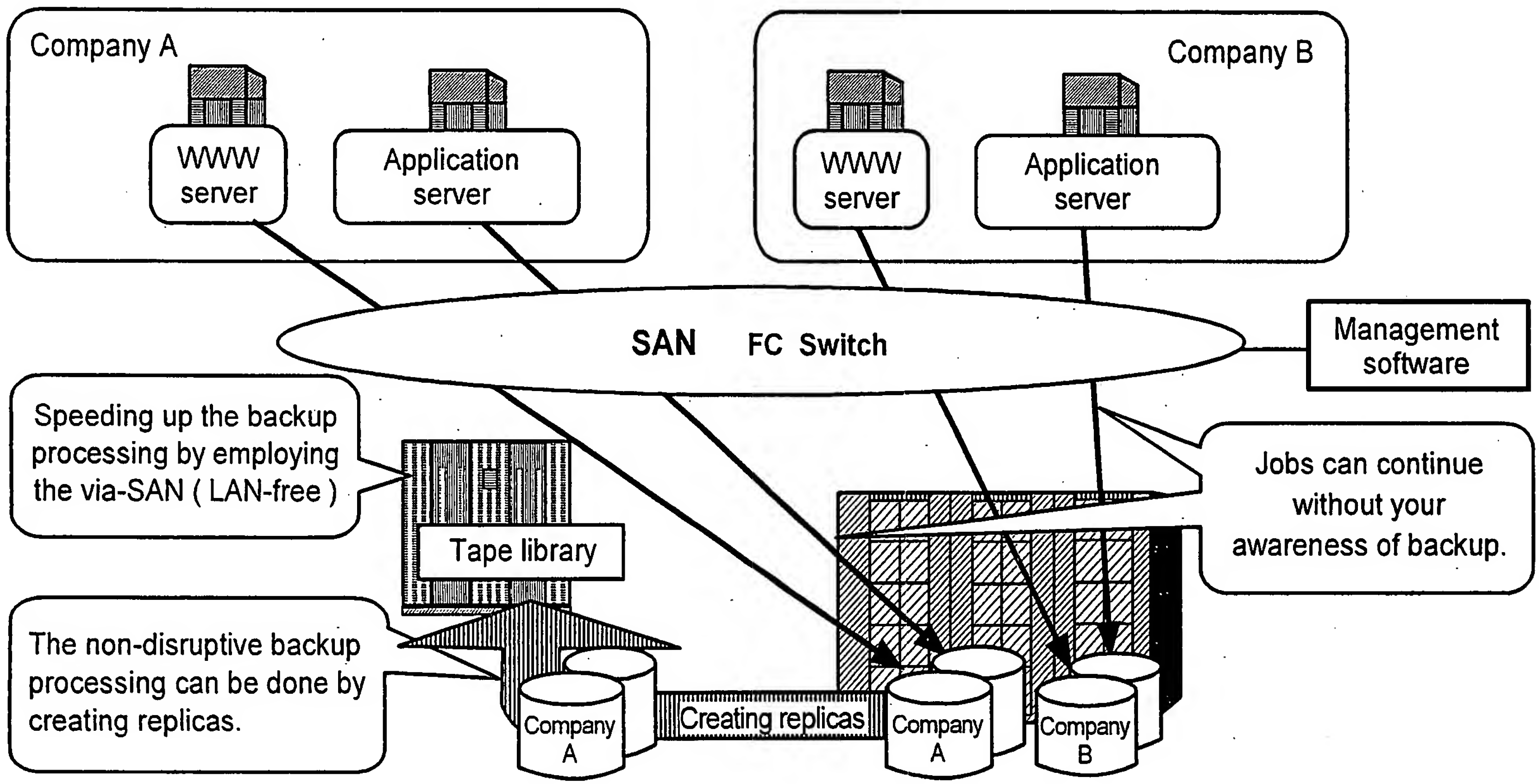

Company A
WWW server
Application server
Company B
WWW server
Application server
SAN FC Switch
Management software
Speeding up the backup processing by employing the via-SAN (LAN-free)
Tape library
Jobs can continue without your awareness of backup.
The non-disruptive backup processing can be done by creating replicas.
Company A
Creating replicas
Company A
Company B

FIG.20

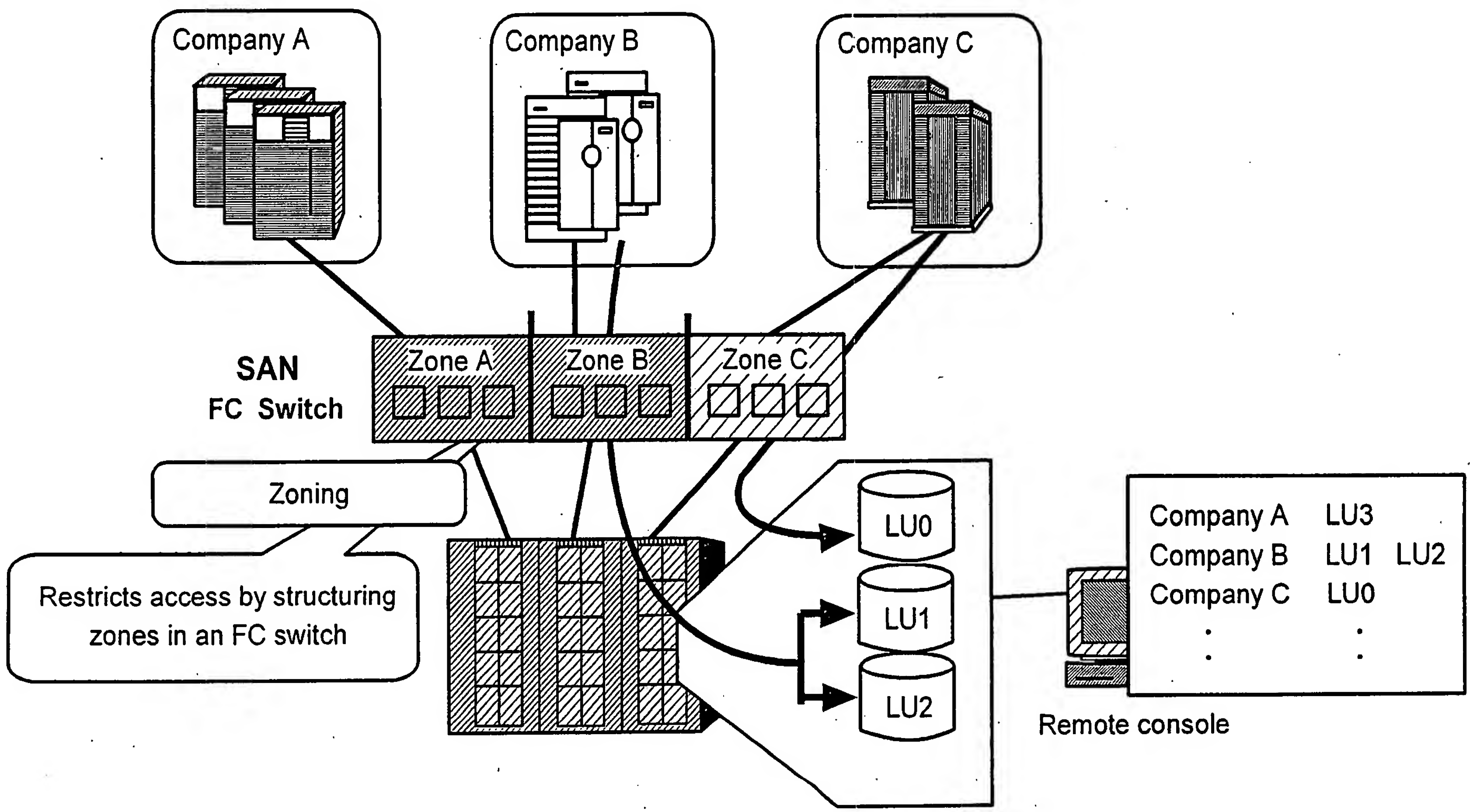
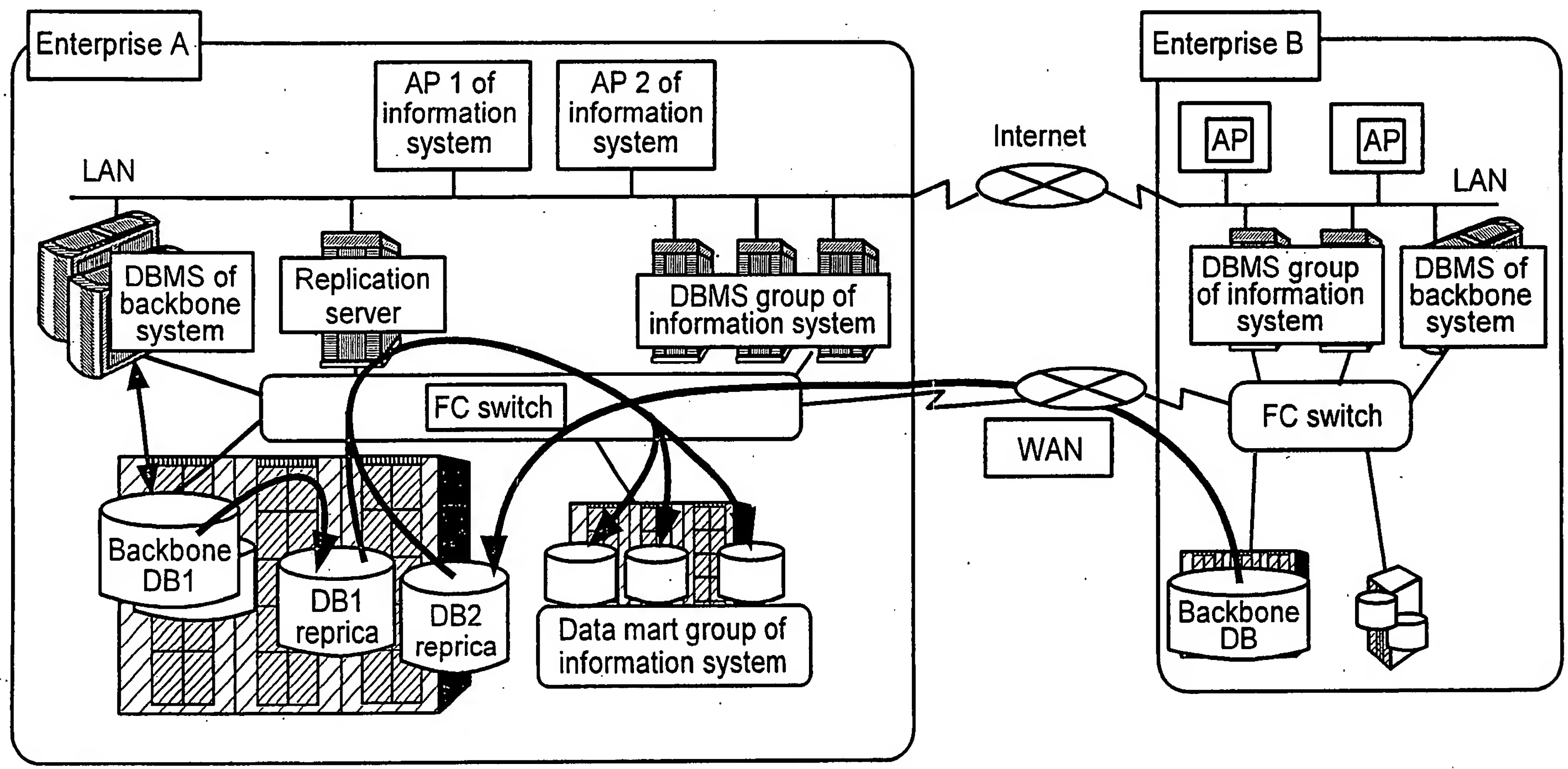

Company A
Company B
Company C
SAN
FC Switch
Zone A
Zone B
Zone C
Zoning
Restricts access by structuring zones in an FC switch
LU0
LU1
LU2
Company A LU3
Company B LU1 LU2
Company C LU0
Remote console

FIG.21


Enterprise A
AP 1 of information system
AP 2 of information system
LAN
DBMS of backbone system
Replication server
DBMS group of information system
FC switch
Backbone DB1
DB1 reprica
DB2 reprica
Data mart group of information system
Internet
WAN
Enterprise B
AP
AP
LAN
DBMS group of information system
DBMS of backbone system
FC switch
Backbone DB